Investigating the Role of the Collagen Protein in *Trichodesmium erythraeum*

A Thesis

Submitted to the Faculty

Of

Drexel University

by

Simara T. Price

in partial fulfillment of the

requirements for the degree

of

Doctor of Philosophy

May 2013

Dedications

To my Aunty Godmother, Elizabeth Wilson, who instilled into me “R.I.F”—Reading is fundamental.

Acknowledgements

I am eternally grateful to my advisor, Dr. Shivanthi Anandan for taking a chance and giving me one. She has given me an opportunity to achieve a level of academic excellence I never knew I had in me and continually had faith in my research and career pursuits.

Thanks to Dr. Joe Bentz for his kindness and always taking the time to hear my concerns and offering advice even when I catch him in the elevator. He is truly the Wizard of Oz in my science world and helped me find my way. A special thank you to Dr. Jennifer Stanford for being one of the best educators I could model myself after and for never wavering in her support of my career goals. To Dr. Bradley Layton, for the opportunity to work on such a unique project. Our road trips to WHOI and UMDNJ to get this project off the ground resulted in conversations that showed me his confidence to think outside of the box and inspired me to do so as well. I am so gracious to Dr. Aleister Saunders for exhibiting the most selfless support to my research. He gave me the opportunity to develop my molecular skills with his students without the slightest hesitation.

I would like to thank Dr. Annette Hynes for opening the doors of her home to some random graduate student from Philadelphia who needed to learn how to culture the most finicky bacterium. She allowed me to shadow her at WHOI and taught me all the tricks a protocol just can't explain. To Dr. Jonathan Zehr for providing a perspective on my project that could have only come from his invaluable expertise in the field.

I am forever indebted to Jonathan Bowen and Lee Dolat for their assistance with the confocal microscopy and helping me to attain the beautiful images that were essential for my research. I would like to thank my lab members, Sondra Schreibman for showing me the ropes when I was a new student, Karl Siegert for keeping me laughing until my stomach hurts and Jordan “Little J” Gallinetti for being the only person to out “Type A” me. Of course, a huge thank you to my unofficial lab partners Suruchi Utreja and Ginnene DiStefano for both intellectual and technical help with my project but most importantly friendship. To my undergrads for sticking with me even though I looked over their shoulder the entire time they even breathed near my project, especially Theresa Weiss and Rhea Thompson, I truly appreciate you. Thank you to my Science Leadership Academy high school students Maya and Brandon, mentoring them was surely more beneficial for me than it was for them.

Chris Telesford, thank you for being the most patient and emotionally supportive man a girl could ask for. Thank you to Adrienne Stevenson M.D. for being such an overachiever that I have no choice but to keep up. Thank you to my grandma for providing me the support to have a fresh start. To Avery for always keeping Ti-Ti smiling and just being silly, it always kept me grounded in reality. To my family for holding me up when I was down, catching me if I was falling and skyping me into every family function I couldn’t physically make. Last, but not least, to my mom for being the ever present rock in my life and the core of all of my success.

Table of Contents

List of Tables

List of Figures

Abstract

Investigating the Role of the Collagen Protein in *Trichodesmium erythraeum*

Simara T. Price

Shivanthi Anandan, Ph.D.

Collagen molecules are structural in nature and primarily found in eukaryotic, multicellular organisms. The collagen protein family is diverse and its membership is continually expanding as new collagen-like molecules are identified. Although previously believed to be limited to eukaryotes, collagen-like proteins have been identified in various prokaryotic organisms. Identification of collagen in unicellular eukaryotes and prokaryotes has opened discussion on the function of these collagens and their role in the emergence of multicellularity. Recently, a collagen-like gene was identified in the marine cyanobacterium, *Trichodesmium erythraeum. T. erythraeum* is a colonial, filamentous cyanobacterium found in tropical and subtropical oligotrophic oceans. Based on the known function of collagen proteins, it was hypothesized that the collagen protein encoded in the *T. erythraeum* genome functioned to provide structural integrity for normal growth and survival of *T. erythraeum.* The collagen protein, named TrpA, was found to be both transcribed and translated at all phases of growth though there is a marked increase in transcription during the logarithmic phase, and a significantly increased amount of protein expressed during both the logarithmic and stationary phase. Immunofluorescent labeling of the TrpA protein revealed that the collagen protein is expressed in the septa between cells along the filament. Additionally, TrpA can be seen localizing to the division septa forming during intercalary cell division. Collagenase treatment of *T. erythraeum* trichomes exhibited fragmentation of the cells along the

filament. Using phylogenetic analysis, structural modeling and circular dichroism, it was confirmed that the protein being expressed belongs to the collagen family, specifically exhibiting characteristics of non-fibrillar collagens. Scanning electron microscopy revealed that TrpA is also being expressed entirely on the surface of the trichomes. While working towards developing a TrpA knockout, it was revealed that *T. erythraeum* is likely naturally transformable. Based on the results found in this study, it is clear that the collagen protein identified in the genome of *T. erythraeum* is involved in maintaining cell-cell adhesion of the cells along the filament. The expression of TrpA on the surface of trichomes indicates that this protein may also be involved in the adherence of trichomes within a colony.

Chapter I. Introduction

I.I. Cyanobacteria

I.I.A. *Significance of Cyanobacteria to Earth's atmosphere*

The composition of Earth's atmosphere from the time of origin to present day can be categorized into three developmental stages (Kasting, 1993). The original atmosphere consisted of carbon dioxide, dinitrogen and reduced sulfur gases. This stage was under fully reducing conditions and completely anoxic. The second stage is described as an oxidizing atmosphere, exhibiting an increase of oxygen. The most recent, and current atmospheric conditions represent the third stage consisting of a fully aerobic atmosphere (Kasting, 1993).

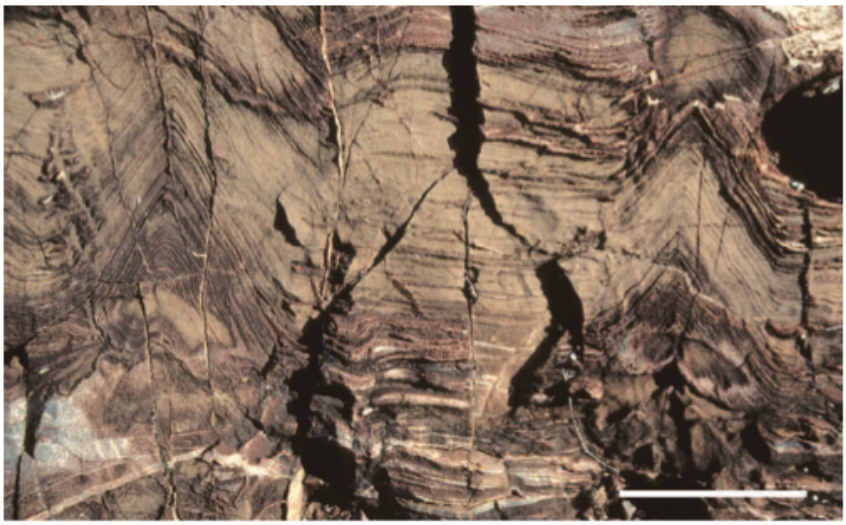


Figure I.1. Stromatolite formation in Strelley Pool

Cyanobacteria have been credited with transforming the prehistoric Earth's atmosphere due to their unique ability to perform oxygenic photosynthesis. Fossils from over two billion year old formations contain the morphotypes of both unicellular and multicellular cyanobacteria coinciding with the "Great Oxygenation Event" (Schirrmeister *et al.*, 2011). Although there is no chemical trace evidence, the morphology of cyanobacteria is unique and carbon isotope ratios support evidence that photosynthesis was involved in the formation of these fossils known as stromatolites (Fig I.1.) (Schirrmeister et al., 2011, Stal, 2012). Though the presence of anoxic microorganisms may have preceded the second stage of the Earth's atmospheric development, the presence of highly organized and complex aerobes on Earth can be attributed to Cyanobacteria.

I.I.B. *Relationship between Cyanobacteria and chloroplasts*

As prokaryotes, cyanobacteria do not have membrane bound organelles in which photosynthesis occurs. However inside cyanobacterium, thylakoid membranes are densely packed in the cytoplasm. Though photosynthesis is exhibited in a number of other bacteria, Cyanobacteria are the only group to exhibit the same type of oxygenic photosynthesis of higher plants (Gupta, 2003). Genomic and biochemical evidence has strongly supported the theory that chloroplasts of higher plants are the result of an endosymbiotic relationship between a cyanobacterial cell and a non-photosynthetic unicellular eukaryote. The cyanobacterium that was taken up by a eukaryote underwent two steps to becoming an organelle. The first would be the transfer of certain genes from its genome to the nucleus of the host cell as well as the evolution of a protein transport system that allowed nuclear encoded proteins to be transported across the double

membrane of the plasmid (Coelho *et al.*, 2013). Even so, the plastids have been shown to contain their own DNA, RNA and ribosomes while exhibiting replication, transcription and translation closely resembling that in bacteria (Raven & Allen, 2003).

I.I.C *Ecology of Cyanobacteria*

Cyanobacteria are a highly diverse group with species varying greatly in both morphology and habitats in which they are found. Upon investigating the cyanobacterial contribution to oxygenating the Earth's atmosphere, the debate ensued as to whether the oldest oxygenic phototroph arose in marine or freshwater environments (Coelho et al., 2013). Evidence shows they likely were originally found in freshwater habitats, however the diversification of Cyanobacteria into their various distinct habitats was essential to the oxygenation of the atmosphere (Coelho et al., 2013).

Whether it is rivers, streams, oceans, glaciers, deserts or hot springs, they are vital primary producers in every environment that they are found in. Aside from their contribution to atmospheric oxygen, these prokaryotes are critical contributors of fixed carbon and even fixed nitrogen in the case of diazotrophs (Capone *et al.*, 1997b). Atmospheric carbon dioxide acts as a carbon source for cyanobacterial carbon fixation. In both freshwater and marine environments, dissolved inorganic carbon (DIC) is an unusable source of carbon for the majority of organisms. Phytoplankton including cyanobacteria have the ability to utilize DIC in the form of carbon dioxide that is then of benefit to the other organisms that are unable to fix inorganic carbon (Su *et al.*, 2012, Capone et al., 1997b). Many cyanobacteria have the ability to fix nitrogen, thus contributing to the availability of fixed nitrogen in their environments. Although not all cyanobacteria have the ability to fix nitrogen, they have been credited with producing the

majority of fixed nitrogen in marine ecosystems (Berman-Frank *et al.*, 2003a). Cyanobacteria have had a continued and profound impact on biogeochemical cycling in their various habitats.

I.I.D. *Morphological characteristics*

Cyanobacteria have a number of unique characteristics that aid in their successful colonization of their preferred ecological niches. As previously mentioned, one of their most notable characteristics is their ability to perform oxygenic photosynthesis. Distinct from higher plants, cyanobacteria also contain phycobilisomes that allow for maximum light harvesting. These accessory pigments extend the range of visible light that cyanobacteria can absorb. Phycobilisomes are a complex of phycobiliproteins that are organized for maximum light transfer with chlorophyll a as the final acceptor (Gantt, 1975). Their absorption peak ranges from 540-650 nm; 540-565 nm for phycoerythrins, 610-640 nm for phycocyanins and about 650 nm for allophycocyanins (Gantt, 1975). They also exhibit complementary chromatic adaptation in which they change the composition of phycobilisomes based on light quantity and quality (MacColl, 1998).

Diazotrophic cyanobacteria have evolved a number of ways to ways to fix nitrogen in the presence of oxygen that would under normal circumstances irreversibly inactivate the nitrogen fixing enzyme nitrogenase. Unicellular cyanobacteria and non-heterocystous cyanobacteria temporally segregate photosynthesis and nitrogen fixation by only exhibiting photosynthesis during the light phase and nitrogen fixation during the dark phase (Zehr, 2011). In the absence of nitrogen a number of filamentous cyanobacteria develop heterocysts. These specialized cells do not perform photosynthesis

and contain nitrogenase thus spatially segregating oxygen and nitrogenase (Adams, 2000).

Due to low affinity between carbon dioxide and ribulose-1,5-biphosphate (RuBisCo) carboxylase, cyanobacteria employ a strategy to prevent the loss of carbon dioxide that that can otherwise easily diffuse through the cell membrane (Espie & Kimber, 2011). Cyanobacterial cells contain inclusion bodies called carboxysomes that contain carbonic anhydrase, which converts bicarbonate to carbon dioxide, and RuBisCo. Because these two enzymes are contained within the carboxysome, bicarbonate is converted to carbon dioxide and fixed by RuBisCo to 3-phosphoglycerate. The colocalization of these two enzymes enhances carbon dioxide fixation (Yeates *et al.*, 2008).

All cyanobacteria contain gas vesicles that allows for buoyancy regulation in response to light. These vesicles are significant for the positioning of the cyanobacterium in the vertical light gradients in aquatic environments (Walsby, 1994). Buoyancy regulation through loss of inflation of the gas vesicles at high irradiance has been highly recorded in a number of cyanobacterial species including *Anabaena*, *Microcystis* and *Trichodesmium* (Walsby, 1994). Positive buoyancy due to these gas vesicles also contributes to cyanobacterial blooms that form on the surface of oceans and lakes (Walsby, 1978).

I.I.E. *Classification of Cyanobacteria*

As one of the largest subgroups of gram-negative prokaryotes, cyanobacteria are morphologically diverse (Stanier & Bazine, 1977). Classification has relied heavily on cell morphology, mode of cell division and presence of nitrogen fixation although more

recent phylogenetic analyses of cyanobacteria have shown that these characteristics are not always an indication of the evolutionary relationship between different species (Sanchez-Baracaldo *et al.*, 2005). When classified according to the governing rules that are used to classify bacteria, cyanobacteria were divided into five sections, Chroococcales (Section I), Pleurocapsales (Section II), Oscillatoriales (Section III), Nostocales (Section IV), Stigonematales (Section V) (Rippka *et al.*, 1979, Waterbury, 2006).

Section I, Chroococales, is composed of unicellular cyanobacteria that reproduce by binary fission or budding. They are the simplest of the five subdivisions, structurally speaking. Distinct from any other prokaryotic group, Section II, Pleurocapsales, reproduce by multiple fission. The vegetative cells are enclosed in a fibrous layer that contains the multiple cells until unicellular structures are released through rupture of the layer. The unicellular structures within the fibrous structure termed baeocytes, then undergo multiple fission and develop a fibrous layer of their own. Cyanobacteria from Section III-V all have a filamentous structure also known as trichomes that often are encapsulated in a sheath. Elongation of trichomes are the result of intercalary cell division in which a transverse wall forms through ingrowth of the peptidoglycan layer followed by the outer membrane. Reproduction is the result of trichome fragmentation or hormogonia released at the ends of trichomes (Rippka et al., 1979, Waterbury, 2006).

Oscillatoriales, Section III, is composed of vegetative cells that have a constant cell diameter and shape within the same individual organism. This characteristic is distinct from the trichomes of Nostocales and Stigonematales in that they both have the capacity for cellular differentiation. Both sections have the ability to form heterocysts in the absence of a fixed nitrogen source, as well as the ability to develop resting cells,

akinetes. The distinguishing factor between the two is the polarity of cell division. Section IV exhibits intercalary cell divisions perpendicular to the long axis of the trichome which results in no true branching. Section V, however, has cell division at various angles that results in multiple widths and lateral branches along the trichome (Rippka et al., 1979, Waterbury, 2006).

I.I.F. *Multicellularity in Cyanobacteria*

Prokaryotes are believed to have developed simple multicellularity around 1.5 billion years before eukaryotes (Schirrmeister et al., 2011). As previously mentioned, cyanobacterial subgroups III-V form multicellular filaments or trichomes. Like the majority of bacterial cells, the cell division protein FtsZ accumulates at the site of cell division directing septum formation in cyanobacteria (Klint et al., 2007, Vishnyakov & Borchsenius, 2007). It has been shown to gather at the leading edges of inward growth in *Nostoc/Anabeana* PCC 7120 during cell division and constrict the cell membrane (Klint et al., 2007). The presence of FtsZ in this filamentous cyanobacteria suggest that there was likely an evolutionary alteration in cell division machinery that resulted in incomplete separation after cell division creating filamentous prokaryotes (Klint et al., 2007)

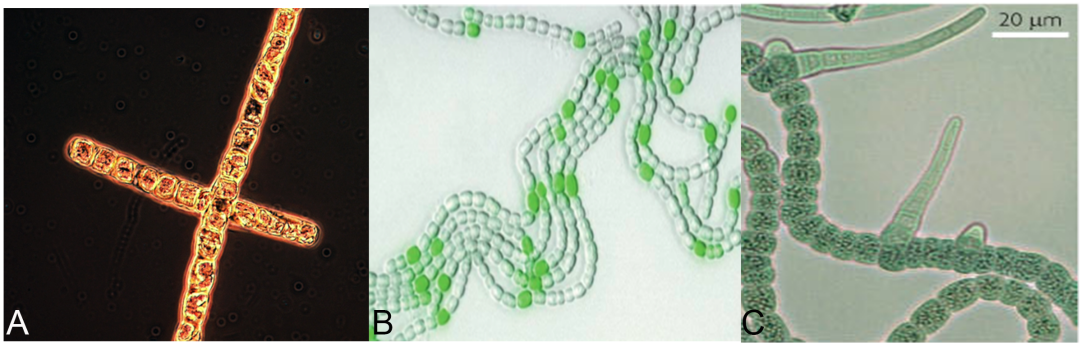


Figure I.2. Multicellular filamentous cyanobacteria from subgroup III-V (A) *Trichodesmium erythraeum* from subgroup III (Hynes 2012), (B) *Anabaena* PCC 7120 from subgroup IV (Golden 2003), (C) *Fischerella* sp. from subgroup V (Flores 2010).

Multicellularity involves cell-cell adhesion, intracellular communication and sometimes cell differentiation (Flores & Herrero, 2010). Recent studies on *Anabaena* and *Fisherella* have shown that the outer membrane is continuous which might help to keep the cells together in the filament. This also suggests that the periplasm is continuous however individual cells have their own cytoplasmic membrane and peptidoglycan layer (Flores & Herrero, 2010). SepJ, a septal protein was recently identified in *Anabaena* PCC 7120 through a transposon mutant that showed increased filament fragmentation. Fluorescence from a SepJ-GFP translational fusion protein was found localized between adjacent vegetative cells clustered in the septa in *Anabeana* PCC 7120 (Flores et al., 2007). The SepJ protein contains a coiled-coil domain that is indicated in protein-protein interactions. Mutant strains with SepJ proteins lacking the coiled coil domain exhibit extensive fragmentation confirming the importance of SepJ in *Anabeana* PCC 7120 (Mariscal et al., 2011). Homologs of SepJ with the coiled-coil domain have been

identified in other filamentous cyanobacteria including *Trichodesmium erythraeum* IMS101, *Nostoc punctiforme* ATCC 29133 and *Lyngbya* PCC 8106 (Mariscal et al., 2011).

I.II. Trichodesmium

I.II.A *Trichodesmium species classification*

Trichodesmium sp. are marine cyanobacteria found in tropical and subtropical oligotrophic oceans (Capone et al., 1997). There are six well defined species of *Trichodesmium*, *T. contortum*, *T. hildebrandtii, T. tenue, T. thiebautii, Katagnymene pelagica/K. spiralis* (*Katagnymene pelagica/K. spiralis* proposed to be the same species), and *T. erythraeum*. *Trichodesmium* sp. have the same photosynthetic system as all cyanobacteria, with the photosynthetic pigment chlorophyll a with light harvesting phycobilisomes. Their primary pigments found in phycobiliproteins are phycoerythrin with absorbance peaks at 490-500nm and 545-565 nm for brown phycourobilin and red phycoerythrobilin (Hynes et al., 2012). Previously, identification for *Trichodesmium* was based on morphological characteristics including cell width, cell length, gas vesicle distribution and colony morphology. However, morphology can be variable based on environmental conditions and many species have more than one morphology type. For example, *T. thiebautii* form both puff and tuft colonies (Fig I.3) (Hynes et al., 2012).

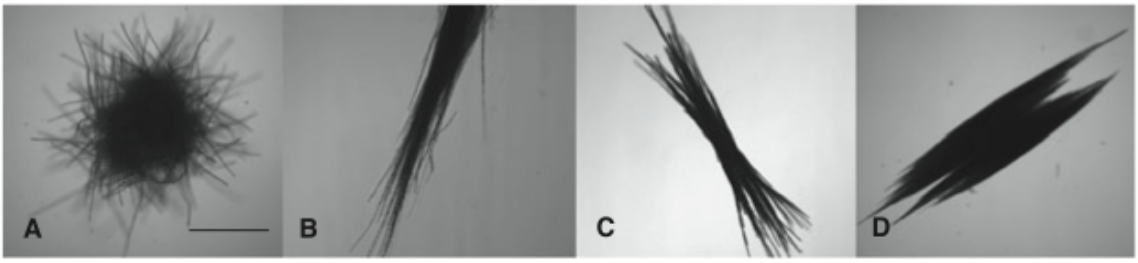


Figure I.3. Colony morphology of *Trichodesmium* spp. Scale bar, 500 μm. (A) puff colony, (B) tuft colony, (C) bowtie colony, (D) raft colony. (Hynes *et al.*, 2012)

Based on the classical morphological groupings, there are two groups, those with peripherally localized gas vesicles including *T. tenue* and *T. erythraeum*, and those with randomly disbursed gas vesicles including *T. thiebautii*, *T. hildebrandtii* and *T. contortum* (Janson *et al.*, 1995). Using genetic techniques, Orcutt et. al. (2002) found that *Trichodesmium* spp. were closely related to one another, yet based on a combination of H1P1 fingerprinting, DGGE analysis of hetR and ITS sequencing, *T. erythraeum* was in a clade of its own while the other *Trichodesmium* species used in the study grouped together in the other clade (Orcutt *et al.*, 2002). Most recently, combining cell morphology, phycobiliprotein spectra, 16s rRNA sequences, Internal Transcribed Spacer (ITS) and hetR, it was determined that there are two distinct clades within *Trichodesmium*: (I) *T. thiebautii*, *T. tenue*, *K. spiralis and T. hildebrantii* and (III) *T. erythraeum* and *T. contortum*. Between clades there is a 97% similarity between the 16s rRNA sequences and within clades there is a 99-100% similarity. While the clades were genetically similar, they also proved to have similar phycobiliprotein composition however the morphological characteristics varied (Hynes et al., 2012).

I.II.B. *Trichodesmium morphology*

Despite their phylogenetic grouping, all *Trichodesmium* spp. have a number of unique characteristics that contribute to their importance in the oligotrophic oceans in which they are found. Interest in *Trichodesmium* was initially sparked after they were identified as a diazotroph in nitrogen limited environments and it has since been credited with ~50% of all nitrogen fixation (Capone et al., 1997). It was found that this non-heterocystous diazotroph fixes nitrogen during the day. This is unique in that all other photosynthetic non-heterocystous diazotrophs fix nitrogen at night because nitrogenase is irreversibly inactivated by oxygen. A number of theories behind mechanisms in which daytime nitrogen fixation occurs in these filaments have been investigated. It had once been suggested that the nitrogenase protein itself had a structural difference that allowed it to function in the presence of oxygen however it was found that there are no unique structural differences (Capone et al., 1997).

The D1 protein of the photosystem II reaction center and the nitrogenase protein both localize to all cells along the trichomes indicating that there is no separation between photosynthesis and nitrogen fixation, however there is a temporal separation of the two processes throughout the photoperiod (Berman-Frank *et al.*, 2001). It was shown that carbon fixation increased in the morning and decreased by midday while nitrogen fixation peaked midday (Berman-Frank 2001). In contrast, some studies have shown that the nitrogenase enzyme is sequestered to a few cells within the trichome that are termed ‘diazocytes’ (Bergman et al., 2012). The frequency of these diazocytes is lower at dawn and increases towards noon, similar to studies that showed there was a temporal separation of oxygen evolution and nitrogen fixation. It is likely a combination of both

temporal and spatial segregation that is preventing the nitrogenase enzyme from being inactivated by oxygen.

Gas vacuoles are an important characteristic for *Trichodesmium* spp. persistence in their marine environments. The vacuoles make up 60-70% of the cell volume and allow for buoyancy regulation (van Baalen & Brown Jr, 1969, Walsby, 1978). About 80% of the trichomes can be found in the top 50m of the eutrophic zone while a significant amount are present at depths greater than 100m. Colonies have the ability to return to the surface faster then single filaments, which is advantageous during periodic mixing (Walsby, 1978). These gas vacuoles are key to *Trichodesmium* spp. ability to form blooms on the oceans surface (Walsby, 1992, Subramaniam *et al.*, 2002). Blooms occur during warm, stable conditions when phosphorus is present and nitrogen is undetectable (Sellner, 1992). Large aggregations of the colonies are huge reservoirs of fixed carbon, nitrogen and phosphorus (Sellner, 1992). Not only do blooms increase primary productivity during for the environment, they also create biological hotspots for a number of microflora found closely associated with the blooms including viruses, bacteria and eukaryotic microbes (Hewson *et al.*, 2009). In fact, some of these viral particles lyse the Trichodesmium cells and cause the extracellular release of organic components into the surrounding water (Hewson et al., 2009).

Buoyancy is regulated by protein and carbohydrate content; by increasing the content buoyancy can be decreased, directly relating vertical distribution to photosynthesis and irradiance (Walsby, 1992). This advantageous adaptation allows for the trichomes to regulate the rate of photosynthesis and exposure to high light. The trichome will float to the surface when irradiance is low which will increase carbon

fixation and carbohydrate production thus decreasing buoyancy and the cycle will continue (Walsby, 1992). At high light irradiance, *Trichodesmium* acclimates to light intensity by down regulating their light harvesting pigments and phycobilins while upregulating light protective carotenoids similar to other cyanobacteria. In addition to buoyancy and photosystem regulation in response to light irradiance, *Trichodesmium* also employs other adaptations to optimize light conditions. Under low light conditions, the cell diameter has been shown to increase thus increasing light absorbance and enhancing the coupling of phycourobilins to PSII rather than PSI (Andresen *et al.*, 2010). The cells also overcompensate for photoinhibitory damage to proteins at high light by enhancing synthesis of the proteins (Andresen et al., 2010).

I.II.C. *Trichodesmium erythraeum* IMS101

At present, the only species of *Trichodesmium* to have its entire genome sequence was *T. erythraeum* IMS101. A large amount of research in *Trichodesmium* has been focused on this strain making it the representative model however, a genetic system as yet to have been developed. The genome is 7.75 Mbp, one of the largest cyanobacterial genomes sequenced yet holds one of the lowest coding percentages (Geer *et al.*, 2010, Bergman et al., 2012). The IMS101 genome is one of the few cyanobacterial genomes that is expanding in size in contrast to the shrinking genome of others. This may suggest that *Trichodesmium* may require flexibility to adapt and cope with the constraints of oligotrophic oceans including horizontal gene transfer and gene duplication to promote genome expansion (Bergman et al., 2012). Recently, a collagen like protein has been identified in the genome of *T. erythraeum* IMS101 (Layton et al., 2008, Orcutt et al., 2002).

I.III. Collagen

I.III.A. *Structure and assembly*

Maintaining structural and mechanical integrity in tissues requires the involvement of connective elements that can withstand stress. Collagen is the most abundant protein in mammals comprising over 30% of protein mass; humans alone contain over 19 distinct collagen types (Ricard-Blum, 2011, Kielty & Grant, 2003). Collagenous domains are comprised of repetitive units of the glycine triplet Gly-X-Y. The amino acid in the X or Y position can be any amino acid, but is most usually proline.

The defining characteristic of collagen molecules is the formation of a left-handed polyproline II helical conformation (Ricard-Blum, 2011). This conformation is a direct result of the primary amino acid sequence. Using fiber diffraction, Ramachandran solved the structure for the collagen triple helix (Ramachandran & Kartha, 1954). It was determined that collagen molecule is composed of three helices (Kadler 2007) (Figure I.4). The glycine in every third position allows for the tight packing of each of the helices (Brodsky & Ramshaw, 1997).

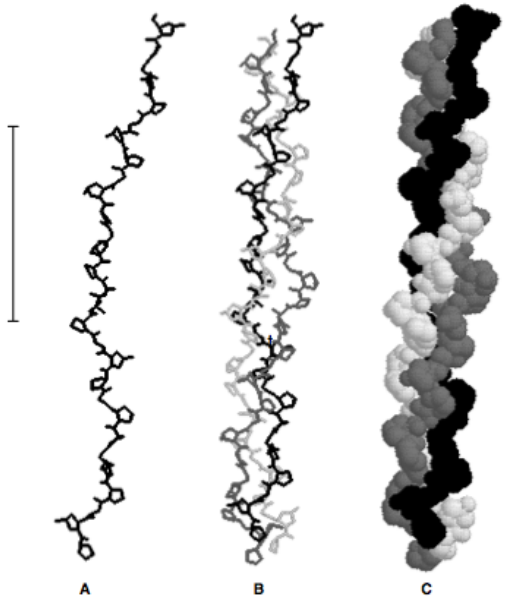


Figure I.4. Structure of a collagen molecule (A) single polypeptide chain (B) trimeric molecule (C) space filling model of trimeric molecule (Kadler 2007).

Collagen polypeptides have non-triple helical domains at their N- and C-termini. These terminal ends are critical in the superstructure the molecules form. The C-terminal end has been implicated in determining which collagen polypeptides will be involved in creating the triple helix (Boudko *et al.*, 2012). Fibril-forming collagens requires the removal of both the N- and C- terminus for fibrillogenesis whereas anchoring fibrils only require the removal of their C-terminus (Kadler et al., 2007).

I.III.B. *Classification within the collagen family*

Collagens were originally classified based on knowledge of the major collagen types. Information was limited to the properties defined by Type I collagen including electron microscopic features, chemical composition and physical properties. Upon the

discovery of new collagens the definition of collagen is a molecule that is comprised of three polypeptide chains that contain repeating Gly-X-Y triplet (Kielty & Grant, 2003).

The classifications of collagens are dictated by the supramolecular structure that the collagen molecules form. They form two main classes fibril-forming and non-fibril-forming collagen (Figure I.6). Fibril-forming collagens all contain long, uninterrupted Gly-X-Y repeat regions that allows these collagen molecules to form characteristic fibrils defined by a banding pattern with a 70 nm periodicity (Gelse *et al.*, 2003). Type I, II, III, V and XI comprise the fibril-forming collagens (Boot-Handford & Tuckwell, 2003). Non-fibril-forming collagens describe a wide variety of supramolecular structures that contain interrupted helical domains. This class includes of basement membrane collagens, anchoring fibrils, fibril associated collagen with interrupted triple helices (FACIT), transmembrane collagens, multiplexins, microfibrillar collagens and hexagonal network-forming collagens (Kadler et al., 2007, Ricard-Blum, 2011).

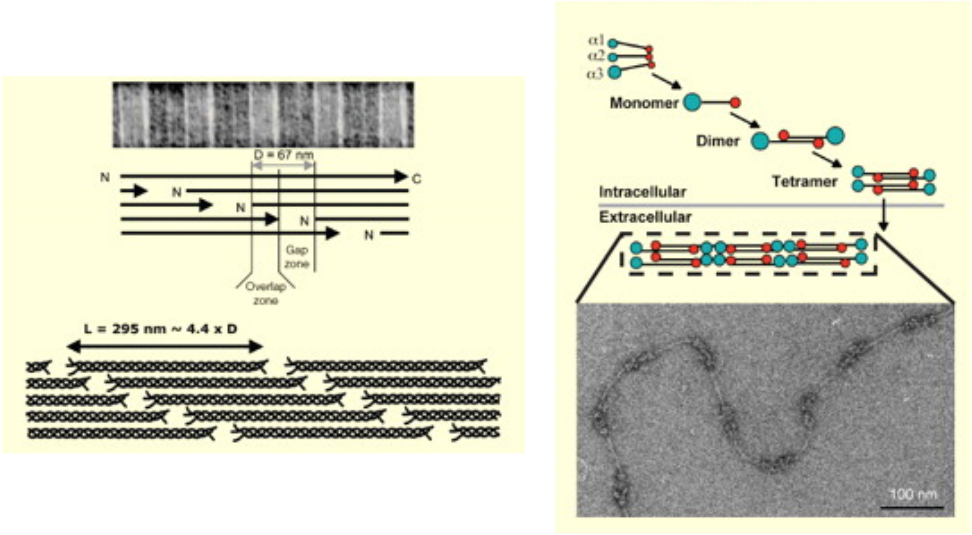


Figure I.4. Fibril and non-fibril forming collagen suprastructures. (A) Fibril forming collagen (B) Non-fibril forming collagen (Kadler 2007)

Previously believed to be limited to vertebrates, collagen and collagen-like molecules have been identified in invertebrates and prokaryotes. Both fibrillar and basement membrane collagens have been described in drosophila, sponges, cnidarian, mussels and hydra (Exposito & Garrone, 1990, Exposito *et al.*, 2008a, Monson *et al.*, 1982). Collagen-like molecules that have been identified and characterized in *Streptoccocus* and *Bacillus* exhibit short collagenous domains (Lukomski *et al.*, 2000, Sylvestre *et al.*, 2002).

I.IV. Rationale for the study

Trichodesmium erythraeum contains the longest collagen sequence identified to date, 1873 amino acids in length. Members of the collagen protein family are involved in maintaining structural integrity and the function of this moleculre in this cyanobacterium was not abundantly clear when this study began. No collagen protein has ever been identified in any other cyanobacteria to date. The aim of this study is to understand the function of this collagen protein in *T. erythraeum*.

Research Hypothesis:

The collagen protein found encoded in the *T. erythraeum* genome has a structural role important for normal growth and survival of this organism.

Based on this hypothesis, the following questions are addressed in this dissertation :

1. Is this collagen gene and protein expressed in *T. erythraeum*?
2. Where is the collagen protein localized in *T. erythraeum* cells?
3. How does the absence of the collagen protein in *T. erythraeum* affect the cell?
4. What classification and type of collagen does *T. erythraeum* collagen most closely relate to?
5. What structure does the *T. erythraeum* collagen peptide chain form?
6. What role does this protein exhibit in *T. erythraeum* cells?

Uncovering the expression and function of this collagen protein in *T. erythraeum* points to the significance of this novel collagen protein in a filamentous prokaryotic species. Specifically, this dissertation reports the involvement of a collagen molecule in cell-cell adhesion in cyanobacteria. Moreover, it demonstrates that this collagen molecule is involved in maintaining filament integrity in *T. erythraeum*.

Chapter 2: Characterization of a Novel Collagen-like Protein in the Cyanobacterium *Trichodesmium erythraeum* IMS101

II.I. Introduction

Collagen is an essential structural protein involved in maintaining the structural and mechanical integrity of vertebrates and other multicellular eukaryotic organisms (Hulmes, 2002). This protein was previously believed to be limited to vertebrates in its distribution, but collagen genes have been identified in unicellular eukaryotes and prokaryotes (Karlstrom *et al.*, 2004, Exposito et al., 2008, King *et al.*, 2008), and are, thus, far more ancient evolutionarily than first thought. These proteins have been implicated in aiding the development of multicellular organisms from unicellular eukaryotic precursor organisms (Abedin & King, 2010). Collagen is the most abundant protein in mammals, and it is responsible for mechanical and structural integrity in every organism from which it has been identified (Nimni *et al.*, 1987) The collagen protein is a major component of various tissues including tendons, cartilage and skin, and also functions in tissue scaffolding and cell adhesion (Kadler et al., 2007).

There are over twenty-eight different types of collagen identified with the most common types of collagen being types I- IV, and VI (Ricard-Blum, 2011). Regardless of the type, collagen proteins contain three polypeptide chains in a triple helical conformation (Shoulders & Raines, 2009). The primary sequence of collagen contains a characteristic Gly-X-Y triglycine repeat region that is essential for the formation of the triple helical structure between three individual collagen peptide strands (Kielty & Grant, 2003). Characterization of collagen is dependent on many factors including the presence

or absence of interruptions in the triglycine repeat region, composition of the triple helix (homo- or heterotrimer), and removal of C- and N- terminal propeptide regions. These factors influence the ability of the triple helix to form supramolecular aggregates including fibrils, other nonfibrillar structures such as Fibril-Associated Collagens (FACITS), transmembrane collagens, network forming collagens, and anchoring fibrils (Kielty & Grant, 2003, Vuorio & de Crombrugghe, 1990). The superquaternary structure formed determines whether the collagen type is classified as fibrillar or non-fibrillar, but in either case the function of the collagen family of proteins is structural in nature (Kielty & Grant, 2003).

Although collagen has been thought of as a strictly eukaryotic protein as mentioned previously, collagen-like proteins have been observed in prokaryotes. The SclC protein in *Streptococcus equi* and SclA and SclB protein in *Streptococcus pyogenes* contain a Gly-X-Y triglycine repeat region (Karlstrom et al., 2004). The exosporium region of the *Bacillus cereus* and *Bacillus anthracis* spores both contain a protein, ExsJ and BclA respectively, that also contain a collagen-like triglycine repeat (Todd *et al.*, 2003, Brahmbhatt *et al.*, 2007). In both cases, it has been suggested that these collagen-like proteins found in these pathogenic bacteria are involved in adherence to cells and tissues, aiding in host cell adherence prior to invasion (Sylvestre *et al.*, 2003, Boydston *et al.*, 2005). Recently, a gene encoding a novel, large collagen-like protein has been discovered in the non-pathogenic, marine cyanobacterium, *Trichodesmium erythraeum* (Layton et al., 2008).

Trichodesmium erythraeum is a filamentous, marine cyanobacterium that has been credited with a large percentage of biological nitrogen fixation in the vast tropical

and sub-tropical oligotrophic oceans (Capone et al., 1997, Bergman et al., 2012). The filaments are referred to as trichomes because of their unbranched, hair-like appearance. Due to its unique role in marine environments as a non-heterocystous nitrogen fixer in the presence of oxygen, significant research has been focused on understanding its physiology and ecological impact (Zehr *et al.*, 2000, El-Shehawy *et al.*, 2003, Berman-Frank et al., 2003, Bergman et al., 2012). *T. erythraeum* encodes a collagen-like protein that contains 558 triglycine repeats (Accession YP_720336.1) (Price & Anandan). This is the longest, uninterrupted collagen triglycine repeat region identified in either eukaryotic or prokaryotic collagen proteins (Layton et al., 2008). This protein has 41% identity to human type II collagen and 45% identity to type I collagen in *Rattus norvegicus*, (Price and Anandan, unpublished data). Atomic Force Microscopy with immunogold labeling identified a protein present on the surface of *T. erythraeum* IMS101 that is cross-reactive with a human anti-collagen antibody (Layton et al., 2008). Phylogenetic analysis suggested that the origin of the collagen protein in *T. erythraeum* occurred during the evolutionary divergence of fibrillar and non-fibrillar collagen (Layton et al., 2008). Based on our knowledge of collagen function in eukaryotic systems in conjunction with the limited data available on the *T. erythraeum* collagen protein, we predict that this protein functions in cell-cell adherence.

To elucidate the function of this collagen protein in *T. erythraeum,* we have utilized immunofluorescence techniques to determine the location of the collagen protein within the cells and the trichome. We have also investigated the pattern of expression of its mRNA and protein to determine if this expression pattern correlated with changes in the growth phases of the *T. erythreaeum* lifecycle. Our data indicate a specific pattern of

collagen protein localization, and is consistent with the known function of the collagen protein in eukaryotes, suggesting that this protein plays a structural role in *T. erythraeum*. Given the similarity of this protein to other triglycine repeat containing proteins, we suggest that this protein be named Triglycine repeat protein A (TrpA), and the gene encoding it *trpA*. We have used this nomenclature for this triglycine repeat protein and gene of *T. erythraeum* in this report.

II.II. Materials and Methods

Bacterial Strains and Culture Conditions

Unialgal *Trichodesmium erythraeum* IMS101 cultures were obtained from Woods Hole Oceanography Institute. Cultures were grown at 26°C with a 14:10 light-dark cycle at 30-40μE m^{-2} s^{-1} in amended seawater (Webb *et al.*, 2001). Cultures were grown in either 250- or 500ml polycarbonate flasks (USA Scientific, Ocala, Florida, USA). Cultures were sampled on days 7, 14, 21, and 28 post inoculation.

RNA Extraction

T. erythraeum samples were collected at the timepoints described above by centrifugation at 3,000 x g for 20 minutes at room temperature. Total RNA was then extracted using the RNeasy Mini Extraction Kit (Qiagen, Valencia, CA, USA), according the manufacturer's instructions. RNA was treated with RQ1 RNase-Free DNase according to manufacturer's instructions (Promega, Madison, WI, USA). RNA concentration was measured by spectrophotometric analysis using a NanoDrop1000 spectrophotometer (Thermo

Scientific, Pittsburg, PA, USA). Samples were then either used immediately as template for cDNA synthesis or stored at -80°C.

Reverse Transcription and cDNA synthesis

50ng of isolated total RNA was used as the template to synthesize cDNA using GoScript™ Reverse Transcription System (Promega, Madison, WI, USA) according to the manufacturer's instructions, using the random primers for cDNA synthesis included in the kit. cDNA concentration was measured using spectrophotometeric analysis. The samples were then either used as template for quantitative PCR (qPCR) analysis or stored at -20°C.

Determination of mRNA levels of the trpA (collagen) and rnpB genes using qPCR

Transcript levels of both the *trpA* and *rnpB* gene were determined using 1µl of the cDNA samples as template for qPCR using the SsoFast™ EvaGreen® Supermix system (BioRad, Hercules, CA, USA) according to manufacturer's instructions. Amplification was performed using a CFX384™ Real-Time System (BioRad) with 20µl of the reaction mixture. Primers were designed to amplify approximately 150-200 base pairs of both the *trpA* and *rnpB* genes (Table 1). Cycling conditions were as follows: 95°C for 30 seconds, 40 cycles of 95°C for 5 seconds, 55°C for 5 seconds followed by a melt curve analysis. Samples from each timepoint were run in quadruplicate replicates. A control without the reverse transcriptase enzyme was used to control for DNA contamination. Expression of the *trpA* gene was normalized to the housekeeping gene *rnpB.* The experiment was repeated 3 times. Amplified products were run on 1.5% agarose gels.

Table II.1. Gene specific primers used for qPCR, 5'-> 3' direction

Genes	Forward primers	Reverse Primers
collagen	CCTATCTGTGCCACCTGGTT	CCGCTAGGGAATTCAACAAA
rnpB	TGGTAACAGGCATCCCAGATAGATA	CGGGTTCTGTTCTCTCAACTCAA

Isolation of crude total protein extracts

T. erythraeum samples from each timepoint were collected by centrifugation at 3,000 x g for 20 minutes. Cells were lysed using Radio-Immunoprecipitation Assay (RIPA) buffer with added 1X Halt Protease and Phosphotase Inhibitor Cocktail and EDTA (Pierce, Rockford, IL, USA). The resulting lysate was centrifuged for 20 minutes at 8,000 x g and the supernatant containing total, cellular proteins was collected and stored at -80°C. Protein concentration was determined using the BCA Protein Assay Kit (Pierce) according to manufacturer's protocol. Spectrophotometric analysis for protein concentration was performed at 562 nm using MultiSkan Spectrum (Thermo Scientific).

Western blotting for collagen protein detection

Protein samples were prepared for SDS-PAGE using NuPage® LDS Sample Buffer (Invitrogen, Grand Island, NY, USA) and RIPA buffer, and heated to 95°C for 10 minutes prior to loading of the gel. Equal concentrations of total protein from each sample were used, and the prepared samples were separated on NuPage Bis-Tris 4-12%

sodium dodecyl sulfate (SDS) polyacrylamide gels in NuPage MES SDS running buffer using the Novex Midi Gel System (Invitrogen). The proteins were electrophoretically transferred to Immobilon-FL PVDF Membrane (Millipore) using the Hoefer Semi Dry apparatus (Hoefer, Hollister, MA, USA) and the membrane blocked in Odyssey Blocking Buffer (Licor, Lincoln, Nebraska, USA) overnight. The membrane was probed with a 1:200 dilution of polyclonal rabbit anti-collagen Type I-V primary antibody (Abcam, Cambridge, MA, USA) for 90 minutes, followed by incubation in 1:5000 dilution of secondary antibody goat anti-rabbit LiCor IR800 (LiCor) in blocking buffer for 60 minutes. The membrane was washed four times at 10 minutes each time with phosphate saline buffer (PBS) with 0.01% Tween20 with gentle agitation in between each antibody incubation. The blot was then imaged using Odyssey infrared imaging system (LiCor), and the intensity of the collagen protein band was measured using Odyssey 3.0 program.

Collagenase treatment of T. erythraeum trichomes

(a) 50ml of culture 21 days post-inoculation was treated with 1 mg/ml of collagenase RMP medium (Sigma, St. Louis, MO, USA) and incubated under normal growth conditions for 5 hours. The cells were then collected and analyzed for the presence of the collagen protein by western blotting, as previously described.

(b) 200µl aliquots of cultures 21 days post-inoculation were incubated with 1mg/ml of collagenase for 5 hours in 96 well plates. The cells were imaged using a TRITC filter on a motorized inverted fluorescence microscope (IX-81, Olympus) driven by SPOT Advanced Modular Imaging Software (SPOT, Michigan, USA). Collagenase treatments were done in duplicate for each culture for three independent trials. *Anabeana* PCC7120

was used as a control in these experiments, and treated in the same way as the *T. erythraeum* trichomes.

Immunofluorescence labeling of the TrpA (collagen) protein and confocal laser scanning microscopy

T. erythraeum trichomes were collected at each time point by filtration onto 40 μm pore, Nylon mesh cell strainer (BioExpress, Kaysville, UT, USA). The cells were fixed, stored, permeablized and probed for the presence of the collagen (TrpA) protein according to the protocol previously described (Taniuchi *et al.*, 2008), using polyclonal anti-collagen Type I-V primary antibody (1:200 dilution) (Abcam 24117) and AMCA conjugated goat anti-rabbit IgG blue fluorescent secondary antibody (1:200) (Abcam 123435). Trichomes were observed after antibody labeling with a confocal laser scanning microscope system (Olympus FV1000) equipped with an inverted microscope (IX-81, Olympus) and Plan Apo N 60x NA 1.42 oil immersion objective (Olympus) driven by FV-10 ASW software. Autofluorescence from the cells was visualized using a TRITC filter and immunofluorescence was visualized using a 50mW DAPI filter. Protein detection was done in triplicate for each timepoint/sample. The immunofluorescent images at 600x total magnification were imported into ImageJ64 (Schneider 2012) and the intensity of the fluorescence was measured in areas between cells where cytosolic pigment autofluorescence was not observed and where collagen immunoreactivity was observed. Intensity was measured by mean gray of pixels in the area selected on the image. Triplicate counts of 20 independent measurements between adjacent cells were done and

a paired T-test was used to determine the significance of protein expression localization between in the junctions between cells.

II.III. Results

We investigated the pattern of expression of a putative collagen protein (TrpA) in *T. erythraeum* using q-PCR and immunoblotting, to determine if its expression was limited to specific phases of growth as a clue to its function in this organism. Transcription of the *trpA* gene was detected by qPCR at each timepoint sampled (Fig. II.1a). Western blotting of total protein extracted from *T. erythraeum* samples indicated the presence of a ~214 kDa protein band at all timepoints (Fig. II.1b). Our data reveal a single, cross-reactive band at ~214 kDa under fully denaturing SDS-PAGE conditions, suggesting that this protein forms a homotrimeric molecule. Our data also indicate that both the *trpA* gene transcript and TrpA protein are expressed at all time points sampled (Fig. II.1). A paired t-test revealed that there is a significant increase in expression of *trpA* mRNA levels at Day 14 as compared to Day 7, Day 21, and Day 28 ($p \leq 0.05$) (Fig. II.2a). TrpA protein expression is significantly increased from Day 7 to Day 14 and Day 21 and significantly decreased from Day 14 to Day 28 ($p \leq 0.05$) (Fig. II.2b).

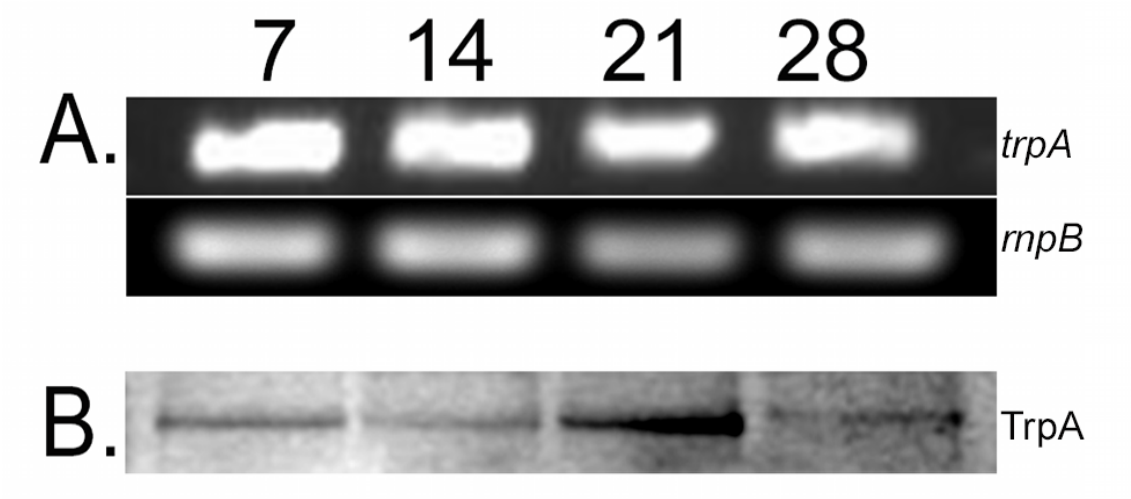


Figure II.1. mRNA and protein expression of the *trpA* gene and TrpA protein at each timepoint sampled (Days 7, 14, 21 and 28 post inoculation). (A) Amplified product from qPCR using *T. erythraeum trpA* (top) and *rnpB* (bottom) gene specific primers. (B) Western blot detecting collagen-like protein TrpA from total protein sample using polyclonal rabbit anti-collagen Type I-V primary antibody.

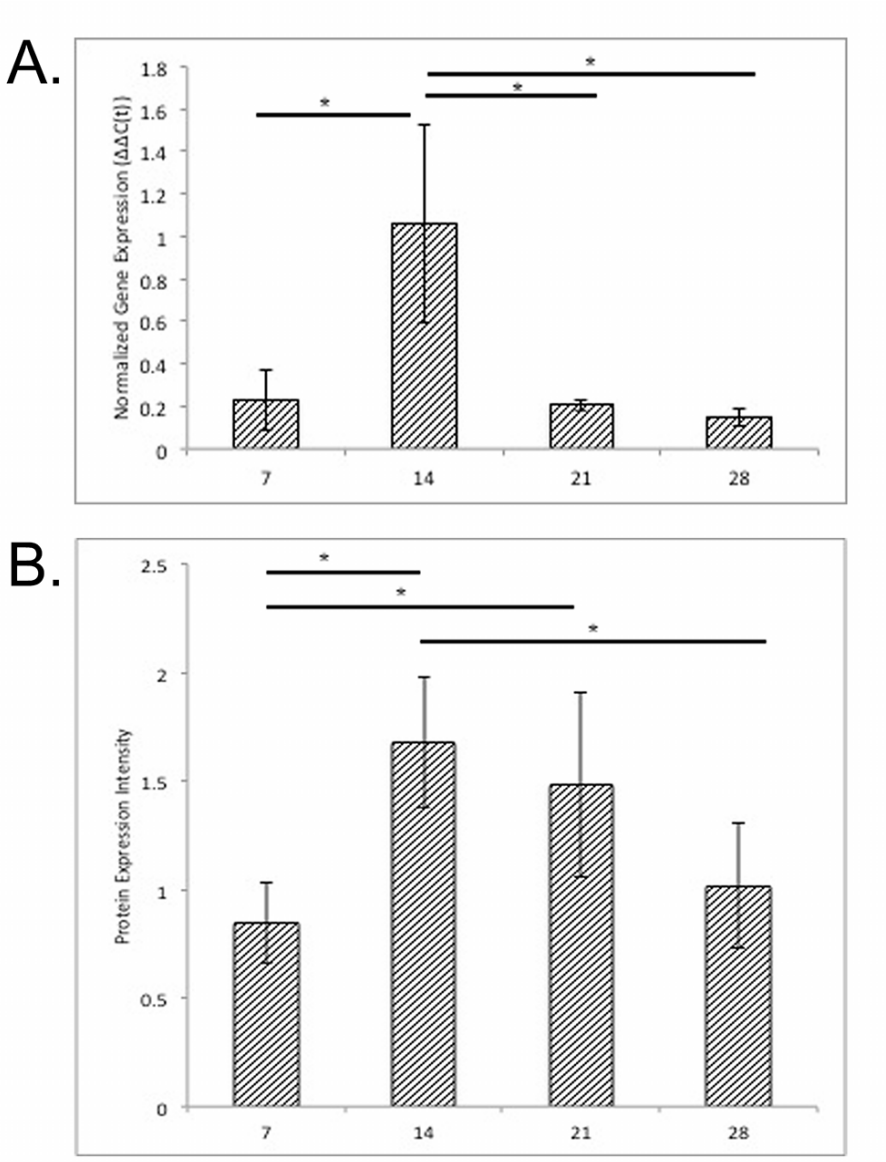


Figure II.2. *trpA* gene and TrpA protein expression at Day 7, Day 14, Day 21 and Day 28 in *Trichodesmium erythraeum*. (A) *trpA* expression is normalized to the housekeeping gene *rnpB*. (B) Expression of TrpA using a polyclonal anti-collagen Type I-IV primary antibody is quantified based on fluorescence intensity (Odyssey, LiCor). These data represent the mean values of at least three independent trials with error bars representing the standard deviation indicated. Asterisk indicates significance ($p \leq 0.05$), paired t-test.

We then used confocal laser scanning microscopy to visualize the location of the TrpA protein in these cells. Cells were labeled with the anti-collagen primary antibody and Aminomethylcoumarin Acetate (AMCA)-conjugated secondary antibody. Fluorescence using the DAPI filter was only observed when both the primary and secondary antibody was used. We specifically used the AMCA-conjugated secondary antibody to distinguish between the fluorescence signal from labeled TrpA, our protein of interest, and the cytosolic autofluorescence from photosynthetic pigments such as phycoerythrin present in *T. erythraeum* that fluoresce in the 495-565 nm range (Subramaniam et al., 1999). At each timepoint, labeled TrpA protein signal was observed within the cells, overlapping with the fluorescence signal from cytosolic photosynthetic pigments. TrpA protein was also specifically localized to the junctions between adjacent cells in a trichome, where cytosolic pigment autofluorescence is negligible (Fig. II.3a,b).

We determined using a paired t-test, that there is a statistically significant difference between the presence of the fluorescence signal of the labeled TrpA protein in the junction of adjacent cells and the absence of the fluorescence signal from the cytosolic photosynthetic pigments ($p<0.0005$). These data indicate that the localization of the TrpA protein to the region between adjacent cells in a trichome is highly specific to this protein. It is also noteworthy that the TrpA protein appears to be localized to the region of the dividing cell where the development of the division septum occurs (Fig II.4, arrows).

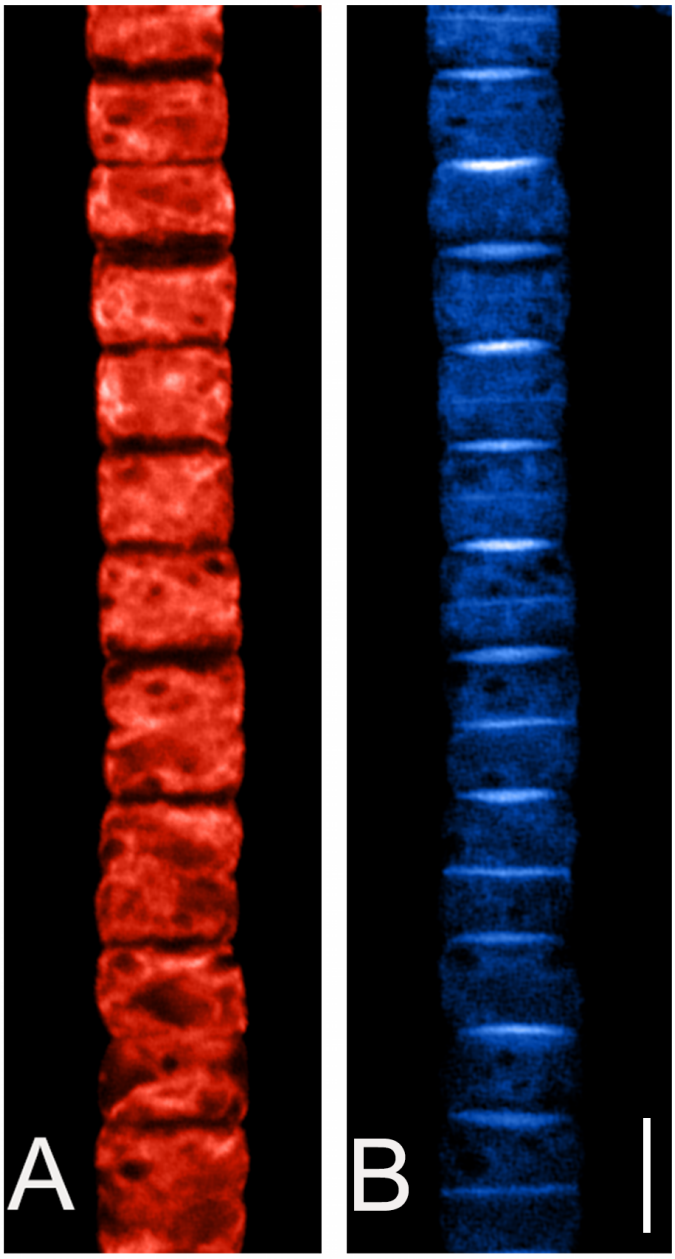


Figure 3. Confocal fluorescence scanning microscopy images showing localization of TrpA protein in *T. erythraeum* IMS101. (A) Autofluorescence of photosynthetic pigments limited to the cells along the trichome using a TRITC filter. (B) Fluorescence under the DAPI filter of the same trichome labeled with an anti-collagen antibody. Localization of the collagen-like TrpA protein can be seen in the junctions between the cells. Scale bar = 5µm

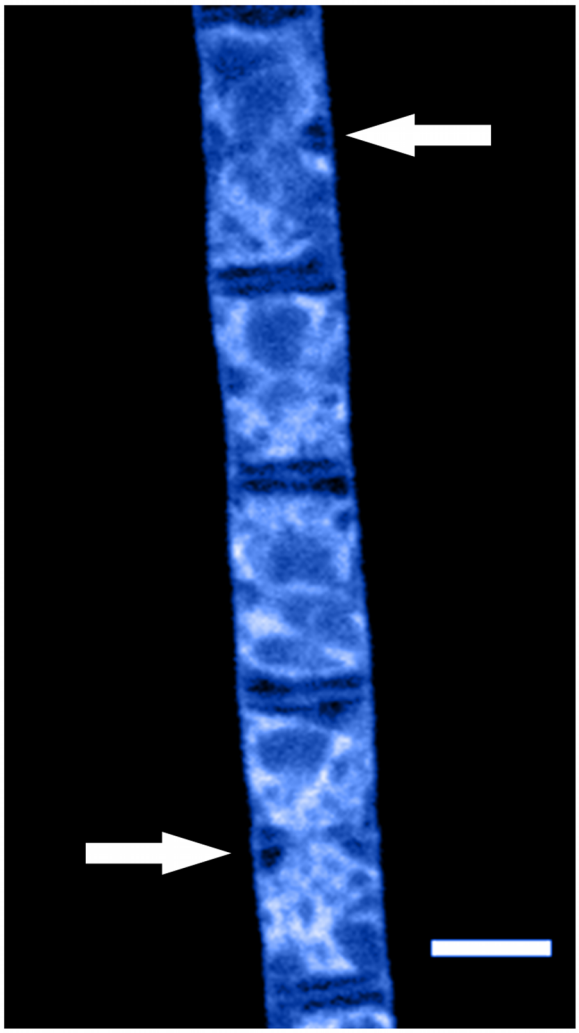


Figure II.4. Localization of the TrpA protein at the division septa (arrows) during intercalary cell division. Scale bar = 5μm

To further confirm that the TrpA protein is located between cells in a trichome, we used the collagenase enzyme from *Clostridium histolyticum*, which specifically degrades collagen protein (Mandl et al 1958). This treatment of *T. erythraeum* cultures prior to sample collection and protein extraction resulted in the complete disappearance of the collagen polyclonal antibody cross-reactive protein band at ~214 kDA,

concomitant with the appearance of smaller antibody cross-reactive bands ranging in size from 41-55 kDA (Fig II.5). These data confirm that the protein located between cells in a trichome is indeed a collagen-type protein. Moreover, *T. erythraeum* cells imaged after collagenase treatment showed increased fragmentation of trichomes (Fig II.6.), suggesting that the TrpA protein promotes adherence of adjacent cells. No filament fragmentation was observed in *Anabeana* PCC7120 under the same treatment, indicating that *Anabaena* PCC 7120 filaments do not possess an externally located collagen-like protein (Data not shown).

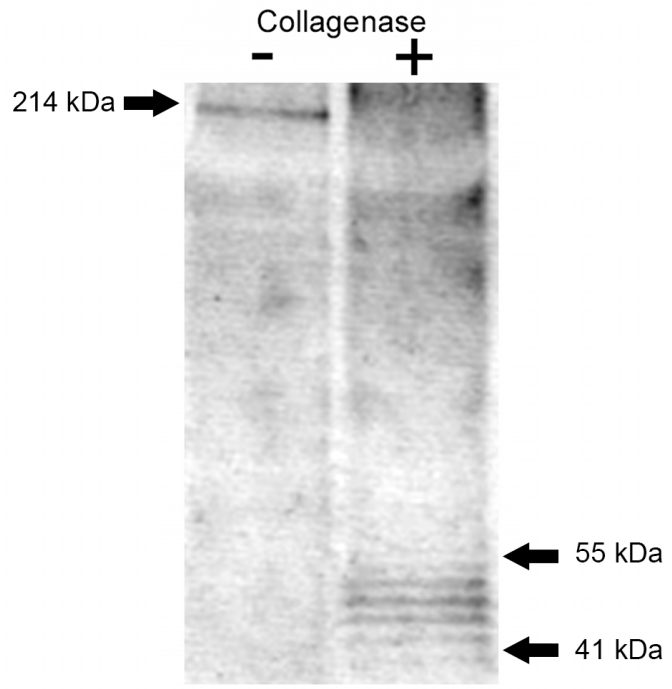


Figure II.5. Collagenase treatment of *Trichodesmium erythraeum* IMS101. SDS-PAGE gel from cultures incubated under normal conditions without collagenase (-col) and with collagenase (+col). Arrow identifies collagen-like TrpA protein

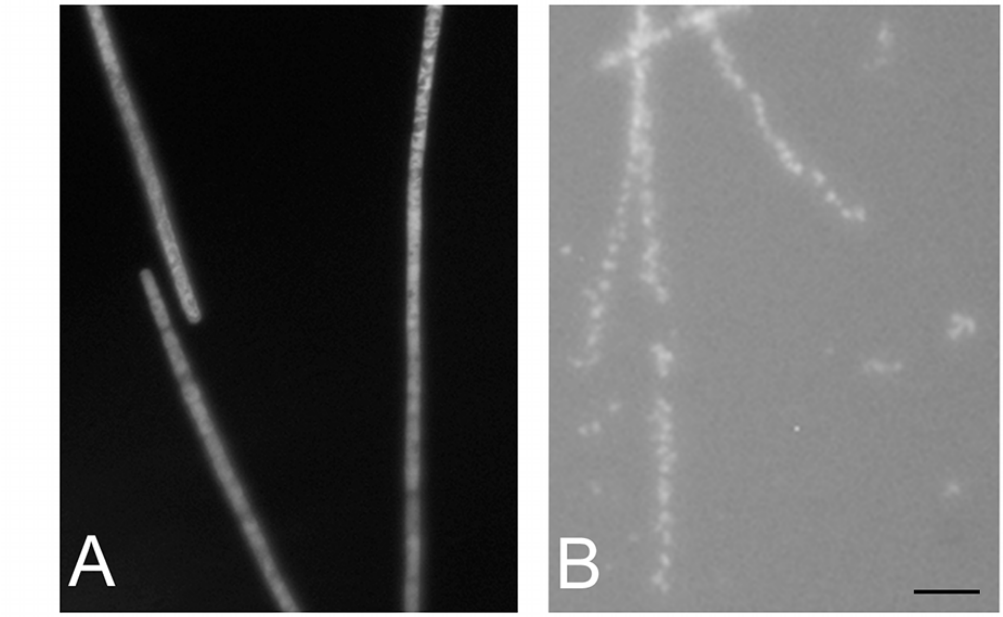


Figure II.6. Effect of collagenase treatment on *Trichodesmium erythraeum* IMS101 trichomes (A) Immediately after addition of collagenase. (B) After 5 hour incubation with collagenase. 200x total magnification

II.IV. Discussion

In this report we analyze the function of a novel collagen-like protein, TrpA, found in the marine cyanobacterium *Trichodesmium erythraeum* IMS101. Although collagen is most commonly associated with multicellularity, collagen-like proteins have recently been identified in bacteria (Rasmussen *et al.*, 2003). In the few prokaryotes where collagen-like proteins have been identified, its function has been implicated in adherence of those bacterial species to other cells or surfaces (Rasmussen *et al.*, 2000, Sylvestre et al., 2003, Karlstrom et al., 2004). However, research on these bacterial collagen-like proteins has been limited.

We have determined that the *trpA* gene is expressed through all phases of growth at both the mRNA and protein levels suggesting that it is important for normal growth of this organism. The significant increase of *trpA* gene expression at Day 14 correlates with active growth in the culture. Being a filamentous cyanobacterium, *T. erythraeum* IMS101 cells maintain contact with adjacent cells in trichomes, which likely explains why expression of this gene and protein was observed at all growth phases under normal culture conditions. We have yet to investigate how different stress conditions affect expression of this gene and protein.

Previous studies that identified known cytosolic proteins in *T. erythraeum* IMS101 such as nitrogenase and rubisco using immunolocalization techniques show that expression of these proteins is limited within the cells of the trichomes (Lin *et al.*, 1998, El-Shehawy et al., 2003). Since TrpA, along with any other protein in bacteria, is translated on cytosolic ribosomes, we expected to detect expression of TrpA within the cell, overlapping with the autofluorescence signal of the known cytosolic photosynthetic pigments. However, our data reveal that TrpA is localized between adjacent cells of the trichome. This specific localization may indicate that TrpA undergoes posttranslational modifications typical of most collagens that render the soluble procollagen protein, insoluble (Canty and Kadler 2005, Kadler 2007). The detection of TrpA in significantly high concentrations between adjacent cells of a trichome reveals a hitherto unknown putative function for this protein in *T. erythraeum* IMS101 cells.

The bacterial enzyme collagenase from *Clostridium histolyticum* specifically cleaves the triple helices of collagen and is often used to liberate cells from tissues and surfaces it is associated with via collagen attachment (Rodbell, 1964, Klock *et al.*, 1996).

Fragmentation of the *T. erythraeum* IMS101 trichomes after treatment with collagenase (Fig. II.6b) indicates that TrpA is involved in joining cells along the filament. Moreover, these data also indicate that it is the triple helical form of the TrpA collagen-like protein in *T. erythraeum* that is localized between adjacent cells of the trichome. While more research must be done to confirm the formation of a triple helix in *T. erythraeum*, collagen-like proteins found in other bacteria have been shown to form typical collagen triple-helix structures (Yu *et al.*, 2010).

T. erythraeum IMS101 is one of the few filamentous marine cyanobacteria in the order Oscillatoriales with its entire genome sequenced, which could explain the singularity of the presence of this collagen-like protein in this organism. However, other cell-cell adhesion proteins have been shown to be present in other filamentous cyanobacteria. Recently, the septum protein, SepJ, has been identified and localized in *Anabeana* PCC 7120 (Flores et al., 2007). Similar to the TrpA protein, the SepJ protein localizes to the septum between each pair of vegetative cells in a filament (Flores et al., 2007, Mariscal et al., 2011). Filamentous growth and elongation in *T. erythraeum* occurs through intercalary cell division, where daughter cells are attached to one another after division (Stanier & Cohen-Bazire, 1977). In filamentous cyanobacteria, the FtsZ protein forms the typical prokaryotic Z-ring arrangement in dividing cells within the filament, but there is incomplete separation of daughter cells after cell division (Klint et al., 2007). This type of growth has been identified as an important evolutionary step towards the development of the first multicellular organisms (Klint et al., 2007). Notably, both the TrpA in *T. erythraeum* and the SepJ protein in *Anabeana* PCC 7120 are observed at the periphery of dividing cells correlating with the site of cell division, a region where Klint

et al. (2007) have shown that the FtsZ protein also accumulates in filamentous cyanobacteria (Flores et al., 2007). Our novel findings provide additional support that there are specific adhesion proteins involved in maintaining multicellularity in filamentous cyanobacteria.

Multicellularity arose multiple independent times during evolution, but a common feature in each of these evolutionary paths is the involvement of proteins that function in cell adhesion (Ruiz-Trillo et al., 2007, Abedin & King, 2010). Our results indicate that TrpA is involved in cell-cell adhesion between adjacent cells in *T. erythraeum* IMS101 trichomes. While the presence of the *trpA* gene could be the result of a horizontal gene transfer, genomic analysis revealed that the *T. erythraeum* IMS101 genome also contains a gene that encodes an integrin β chain-like protein (Accession: YP_721619.1) (Layton et al 2008) (Price and Anandan, unpublished data), indicating that TrpA could be involved in forming a structure that resembles an extracellular matrix. Bacterial cells are known to secrete MSCRAMMs (microbial surface components recognizing adhesive matrix molecules) that can interact with collagen resulting in adhesion of cells (Chagnot *et al.*, 2012). The precise mechanism by which TrpA supports cell adhesion in *T. erythraeum* will be further investigated in this laboratory.

This chapter has been submitted and accepted by the Journal of Phycology.

Chapter III: Uncovering the Structure and Classification of TrpA

III.I Introduction

Collagens are the most abundant structural protein found in vertebrates. There are currently over 28 different types of collagen (Shoulders & Raines, 2009). Despite this variety, they all have a common feature, a triglycine repeat region in the primary structure leading to the formation of a triple helical tertiarty structure. This characteristic sequence consists of $(Gly\text{-}Xaa\text{-}Yaa)_n$ where X is most commonly proline and Y can be any amino acid but is usually hydroxyproline. The high proline content of collagen chains result in a polyproline II (PPII) left-handed helical structure (Brodsky & Ramshaw, 1997). The triple helical collagen molecule can be composed of three identical (homotrimer) or similar (heterotrimer) collagen chains (Hulmes, 2002).

Interruptions in the triglycine repeat regions introduce a perturbation into the structure of the polypeptide that prevents the perfect triple helical structure. It is important to note that interruptions are not representative of a mutation that negatively affects the function of the collagen structure that it forms, it is only when the interruption is a mutation in a fibril forming collagen that pathology is observed (Bella *et al.*, 2006). These interruptions introduce flexibility into the structure of the triple helix that are critical to the supermolecular structures that nonfibrillar collagens form (Hwang & Brodsky, 2012). These include network-forming collagens, beaded filament forming collagen, anchoring fibrils, fibril-associated collagens, endostatin-producing collagen, and transmembrane collagens (Kadler et al., 2007). For example, the anchoring fibril formed from Type VII collagen contains a nonhelical domain in the center of its sequence

that allow for the fibers to form an arch structure as opposed to a straight fibril (Burgeson, 1993, Bächinger *et al.*, 1990). The arch occurs where the nonhelical domain is found in the polypeptide (Bächinger et al., 1990). These archs surround and entrap collagen fibrils securing them in the basement membrane (Burgeson, 1993).

Previously believed to be limited to vertebrates, collagen and collagen-like proteins have been identified in invertebrates and prokaryotes. The presence of both fibrillar and nonfibrillar collagens in invertebrates including sponge, worm and abalone has provided evidence that a hypothetical ancestor of metazoans contained collagen (Exposito *et al.*, 2002). Collagenous domains in prokaryotes such as the BclA protein in *Bacillus anthracis* and SclA protein in *Streptococcus pyogenes*, could also indicate that evolution of the collagen protein predates the divergence of prokaryotes and eukaryotes and aided in the formation of multicellular organisms. However, horizontal gene transfer from eukaryotes into prokaryotes is also a possible pathway for the presence of this eukaryotic protein in prokaryotes (Exposito et al., 2002).

The function of collagen-like proteins in prokaryotes is not abundantly clear, however they have been indicated in being involved in the adhesion of bacterial cells to their host human cells (Sylvestre et al., 2003). *S. pyogenes*, a bacterial pathogen involved in a wide range of human diseases by attaching to cell surfaces, contains the collagen-like proteins, SclA and SclB that are found on the outer surface of its cells (Rasmussen et al., 2000, Whatmore, 2001). The BclA protein in *Bacillus anthracis*, the causative agent of anthrax, is found on the surface of the exosporium layer of the spore. These spores interact and bind with cells within mammalian hosts and its possible that the BclA protein is involved in this attachment (Boydston et al., 2005).

Recently, a collagen-like protein, TrpA, was identified in the genome of the marine cyanobacterium, *Trichodesmium erythraeum*. It has been shown that TrpA is expressed by *T. erythraeum* cells (Price and Anandan). In addition it is localized to the extracellular space between adjacent cells along the filament in conjunction with other proteins involved in maintaining attachment of the cells along the filament (Price and Anandan) (Mariscal et al., 2011, Flores et al., 2007). In this study, we examined used phylogenetic analysis to determine the relatedness of TrpA to other collagen proteins. Model synthetic peptides were used to determine if the structure of the TrpA protein could forms PPII helices. Scanning Electron Microscopy (SEM) with silver enhancement was used to locate TrpA on the surface of *T. erythraeum* suggesting a secondary function of TrpA. We suggest that TrpA be classified as a nonfibrillar collagen based on the data presented in this study.

III.II Materials and Methods

III.II.A Bioinformatics

Database searching

80 amino acid sequences for fibrillar and nonfibrillar vertebrate and invertebrate collagen chains were obtained from the National Center for Biotechnology Information (NCBI) (Geer et al., 2010). The same species were only used more than once for the same type of collagen if it was a different chain type (ex. *Homo sapien* Type I Chain 1 and *Homo sapien* Type I Chain 2). In addition, TrpA, the collagen-like protein found in

Trichodesmium erythraeum IMS101 was obtained from NCBI. The accession numbers and species are compiled in Table III.1.

Alignment and Analysis

All 80 fibrillar and non-fibrillar collagens and the TrpA collagen-like protein were aligned using Seaview (Gouy *et al.*, 2010). After the sequences were aligned, phylogenetic trees were computed by maximum likelihood (ML) driving program PhyML 3.0 (Guindon *et al.*, 2005). The maximum likelihood tree was midpoint rooted using the Interactive Tree of Life (iTOL) due to the absence of a good outgroup (Letunic & Bork, 2011).

Table III.1. Vertebrate, Invertebrate and Prokaryotic Collagens. Accession numbers, organism, collagen type and chain used for phylogenetic analysis.

	Organism	Collagen Type	Chain	Accession
1	*Trichodesmium erythraeum*			ABG49863.1
2	*Homo Sapien*	I	1	P02452.5
3	*Rattus norvegicus*	I	1	P02454.5
4	*Danio rerio*	I	1	AAH63249.1
5	*Danio rerio*	I	3	AAH58045.1
6	*Gallus gallus*	I	2	P02467.2
7	*Canis lupus*	I	1	NP_001003090.1
8	*Paralichthys olivaceus*	I	1	BAD77968.1
9	*Oncorhynchus mykiss*	I	1	BAB55661.1
10	*Bos taurus*	I	1	P02453.3
11	*Gallus gallus*	I	3	NP_990711.2
12	*Homo Sapien*	III	1	P02461.4
13	*Mus musculus*	III	1	AAH52398.1
14	*Rattus norvegicus*	III	1	NP_114474.1
15	*Sus scrofa*	III	1	NP_001230226.1
16	*Bos taurus*	III	1	NP_001070299.1
17	*Oncorhynchus mykiss*	III	1	NP_001117678.1
18	*Bos taurus*	IV	1	NP_001159983.1
19	*Oryzias latipes*	IV	1	NP_001170943.1
20	*Rattus norvegicus*	IV	1	NP_001128481.1
21	*Macaca mulatta*	IV	1	EHH29119.1
22	*Ascaris suum*	IV	1	ADY40208.1
23	*Mus musculus*	IV	1	AAA50292.1
24	*Gallus gallus*	IV	1	AAF44681.1
25	*Homo Sapien*	IV	1	AAA53098.1
26	*Gallus gallus*	IV	2	NP_001155862.1
27	*Strongylocentrotus purpuratus*	IV	2	NP_999676.1
28	*Brugia malayi*	IV	2	AAC46611.1
29	*Bos taurus*	IV	2	DAA23684.1
30	*Drosophila melanogaster*	IV	2	AAB64082.1
31	*Homo Sapien*	IV	2	CAH72050.2
32	*Mus musculus*	V	1	AAI38450.1
33	*Sus scrofa*	V	1	NP_001014971.1
34	*Gallus gallus*	V	1	NP_990121.1
35	*Heterocephalus glaber*	V	1	EHB05599.1
36	*Homo Sapien*	V	1	BAG48312.1
37	*Bos taurus*	V	3	DAA27964.1
38	*Danio rerio*	V	1	ADG36303.1
39	*Rattus norvegicus*	V	1	AAF76433.1
40	*Mus musculus*	V	2	AAH55077.1
41	*Rattus norvegicus*	V	2	NP_445940.1
42	*Sus scrofa*	V	2	NP_001098759.1
43	*Rattus norvegicus*	V	2	EDL99127.1
44	*Homo Sapien*	V	3	EAW84061.1
45	*Cricetulus griseus*	VII		AAA36968.1
46	*Rattus norvegicus*	VII	1	NP_001100328.1
47	*Canis lupus*	VII	1	NP_001002980.1
48	*Microtus ochrogaster*	VII	1	ADN07507.1
49	*Homo Sapien*	VII	1	AAA75438.1
50	*Bos taurus*	VII	1	DAA17035.1
51	*Mus musculus*	VII	1	EDL21325.1
52	*Rana catesbeiana*	I	2	O42350.1
53	*Paralichthys olivaceus*	I	2	BAD77969.1
54	*Oncorhynchus mykiss*	I	2	BAB55663.1
55	*Oncorhynchus mykiss*	I	3	BAB55662.1
56	*Mus musculus*	I	1	P11087.4
57	*Riftia pachyptila*	Fibrillar		AAB24972.1
58	*Alvinella pompejana*	Fibrillar		AAC35289.2
59	*Mytilus edulis*	Fibrillar		AAB96638.1
60	*Tetraodon nigroviridis*	Fibrillar		CAG02093.1
61	*Ephydatia muelleri*	Fibrillar		1920343A
62	*Bacillus cereus*	Collagen-like		NP_981161.1
63	*Streptococcus equi*	Collagen-like		AAQ91575.1
64	*White spot syndrome virus*	Collagen-like		AAK77699.1
65	*Anopheles gambiae*	Collagen-like		XP_320884.4
66	*Ephydatia muelleri*	Fibrillar	1	P18856.2
67	*Amphimedon queenslandica*	Fibrillar	7	CAQ63563.1
68	*Amphimedon queenslandica*	Fibrillar	6	CAQ63562.1
69	*Amphimedon queenslandica*	Fibrillar	5	CAQ63561.1
70	*Amphimedon queenslandica*	Fibrillar	1	CAQ63559.1
71	*Danio rerio*	II	1	AAI63934.1
72	*Mus musculus*	II	1	NP_001106987.2
73	*Homo Sapien*	II	1	P02458.3
74	*Equusca ballus*	II	1	NP_001075233.1
75	*Oryctolagus cuniculus*	II	1	NP_001182600.1
76	*Macaca mulatta*	II	1	EHH20668.1
77	*Xenopus Laevis*	II	1	NP_001081260.1
78	*Homo Sapien*	II	1	AAC41772.1
79	*Gallus gallus*	II	1	AAO33039.2
80	*Bos taurus*	II	1	NP_001001135.2
81	*Rattus norvegicus*	II	1	NP_037061.1

III.II.B Structural Studies

Peptides

A peptide 13 amino acids in length was constructed using the amino acid sequence for TrpA including the interruption in the sequence; Gly-Pro-Ala-Gly-Pro-Ala-Gly-Gly-Pro-Ala-Gly-Pro-Ala. Two additional peptides 12 amino acids in length were constructed; a mutant of the TrpA sequence without the interrupting glycine $(Gly\text{-}Pro\text{-}Ala)_4$ and a control peptide $(Gly\text{-}Pro\text{-}Pro)_4$. The peptides were acetylated at the N-Terminal and amidated at the C-Terminal and synthesized by GenScript with >98% purity (GenScript, Piscataway, NJ).

Circular Dichroism

The structural conformations of the peptides were determined using circular dichroism (CD). Peptides were solubilized in (85/15) methanol/acetic acid solvent (Pêcher *et al.*, 2009) to a final concentration of 3μM. Temperature-dependent UV-CD spectra were obtained as described in Toal et. al 2011 (Toal *et al.*, 2011). Changes in peak value were calculated by dividing the difference in the absolute peak value by the maximum absolute peak.

III.II.C Culture Conditions

T. erythraeum IMS101 cells were grown under the same conditions as described in chapter 1.

III.II.D Scanning Electron Microscopy

Immunohistochemistry

T. erythraeum IMS101 filaments were collected through filtration onto a sterile 40um pore, nylon mesh cell strainers (BioExpress, Kaysville, USA). The cells were incubated with polyclonal anti-collagen Type I-V primary antibody (1:200 dilution) (Abcam 24117) at room temperature for 30 minutes followed by incubation with 5nm colloidal gold anti –rabbit IgG secondary antibody (Sigma-Aldrich, St. Louis, USA) (G7277) for 30 minutes. Cells were rinsed with RMP media 4 times for 30 seconds each rinse after both antibody incubations. The labeled cells were treated for silver enhancement according to the manufacturers protocol (Sigma, St. Louis, USA). Cells were then fixed in 2.5% gluteraldehyde for 1.5 hours, filtered onto 1.0 um Nuclepore Polycarbonate filters (Fisher Scientific,) and post-fixed in 0.1% osmium tetraoxide for 10 minutes. Cells were rinsed in 0.1 M 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) 4 times for 30 seconds each rinse after incubation in each fixative.

Sample Preparation and Imaging

Fixed cells were prepared for Scanning Electron Microscopy (SEM) imagining as previously described (Marga *et al.*, 2005) except the sample was coated with carbon. Samples were examined in a Zeiss Supra50VP field-emission scanning electron microscope at 2kV and 7.5kV using the SE2 or In Lens secondary electron detector (Zeiss). X-ray microanalysis was used to identify the elemental constituents of the sample. Peaks correspond to specific X-ray lines corresponding with specific elements (Oxford Instruments, Wycombe, UK).

III.II.E TrpA Modeling

Protein Structure Prediction

The I-Tasser server was used to create a three-dimensional structure of a portion of the TrpA protein (Zhang, 2008). The 13 amino acid sequence (Residue 723 to Residue 735) used to create the peptide for CD from the TrpA sequence containing the glycine interruption was input into the I-TASSER server. A second, longer (31 amino acids) sequence from the TrpA protein containing the same glycine interruption (Residue 711 to Residue 741) was also input into the I-TASSER server. Five models were created for each sequence that was input into the server, the first model was chosen as the structural model in both cases because the modeling quality for lower-rank models was much weaker than that of the first model given (Zhang, 2008). Predicted protein structure models were viewed as provided by I-TASSER and/or in DeepView (Guex & Peitsch, 1997).

Homology Modeling

Based on the structure that I-TASSER ranked as the top template for protein prediction, Protein Database Code (PDB) code and chain identifier 2klwA was used as the template for the Automatic Modelling Mode in the SWISS-MODEL Workspace (Schwede *et al.*, 2003). The target sequence that was input was the 31 amino acid sequence that was also used for the I-TASSER protein structure prediction. The PDB code 2klw encodes the structure of a synthetically constructed collagen heterotrimer that was resolved by solution Nuclear Magnetic Resonance, 2klwA is the chain identifier for Chain A of which is 1 of 3 chains. The sequence identity between the template and target sequence was

64.284%, which is greater than the rule of thumb of 50% minimum making it an appropriate submission for the 'Automatic Modelling Mode' (Arnold *et al.*, 2006). Visual inspection of the model and template structure was done using DeepView (Guex & Peitsch, 1997).

All peptide sequences used in this study for peptide studies and modeling are compiled in Table 2.

Table III.2. Peptide sequences used for circular dichroism and protein prediction/homology modeling. Protein Database (PDB) code 2klwA sequence is represented by $(GPK)_{10}$

Sequence	Length
$(GPP)_4$	12
$(GPA)_4$	12
$(GPA)_2G(GPA)_2$	13
$(GPA)_6G(GPA)_4$	29
$(GPK)_{10}$	30

III.III Results

Bioinformatics

Multiple alignments of fibrillar and nonfibrillar collagen protein sequences were performed to explore whether TrpA clustered with fibrillar or nonfibrillar collagens. Phylogenetic distribution of fibril-forming and nonfibril-forming collagens supported the classification of each collagen type. All of the fibril forming collagens clustered together as did all of the non-fibrillar collagens with the exception of invertebrate fibrillar

collagens that were found clustered with the nonfibril forming collagens (Figure III.1). TrpA was found within the portion of the phylogenetic tree that contained the non-fibril forming collagens. With that section, it was most closely associated with invertebrate fibrillar collagens and the two prokaryotic collagen-like sequences that were used in this analysis (Figure III.1).

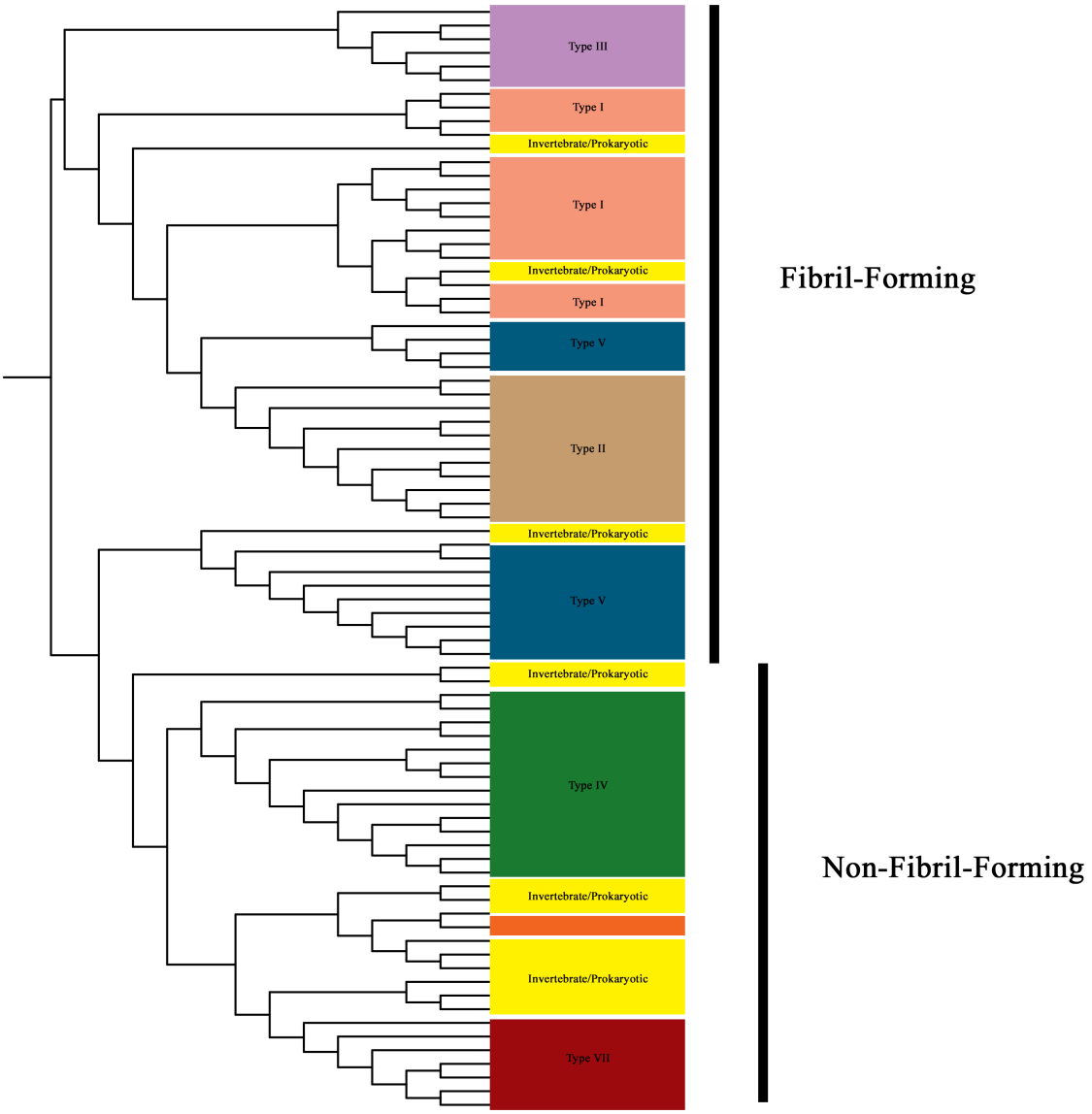


Figure III.1. Maximum likelihood phylogenetic tree created from a multiple alignment of the collagen proteins listed in Table 1. The different collagen types, I-V, VII and Invertebrate/Prokaryotic, are indicated by the following color labels Type I (pink), Type II (tan), Type III (purple), Type IV (green), Type V (blue), Type VII (red), Invertebrate/Prokaryotic (yellow) and *T. erythraeum* collagen (orange). Collagen classification, fibril-forming and non-fibril forming, is indicated on the right hand side of the tree based on the classification of the majority of collagens in that region. The tree is midpoint rooted.

Circular Dichroism

UV-CD spectra were obtained for each peptide $(GPP)_4$, $(GPA)_4$ and $(GPA)_2G(GPA)_2$ between temperatures 5 and 95°C (Fig III.2). All three peptides exhibit polyproline II (PPII) preference with a peak at 215 nm and a negative maximum at 195 nm. An isodichroic point is apparent at 205 nm of each spectrum representing the two states of each peptide characteristic of PPII peptides. $(GPP)_4$ exhibited the most stable PPII conformation with the highest positive peak at 215 nm. When the proline in the Y $(GP\underline{P})_4$ position is changed to an alanine $(GP\underline{A})_4$ the PPII content is decreased by ½. The presence of a glycine insertion, $(GPA)_2G(GPA)_2$, exhibited a further decrease in the PPII content. The decrease with the glycine insertion was by ½ in comparison to the $(GPA)_4$, and a factor of ¾ in comparison to the $(GPP)_4$.

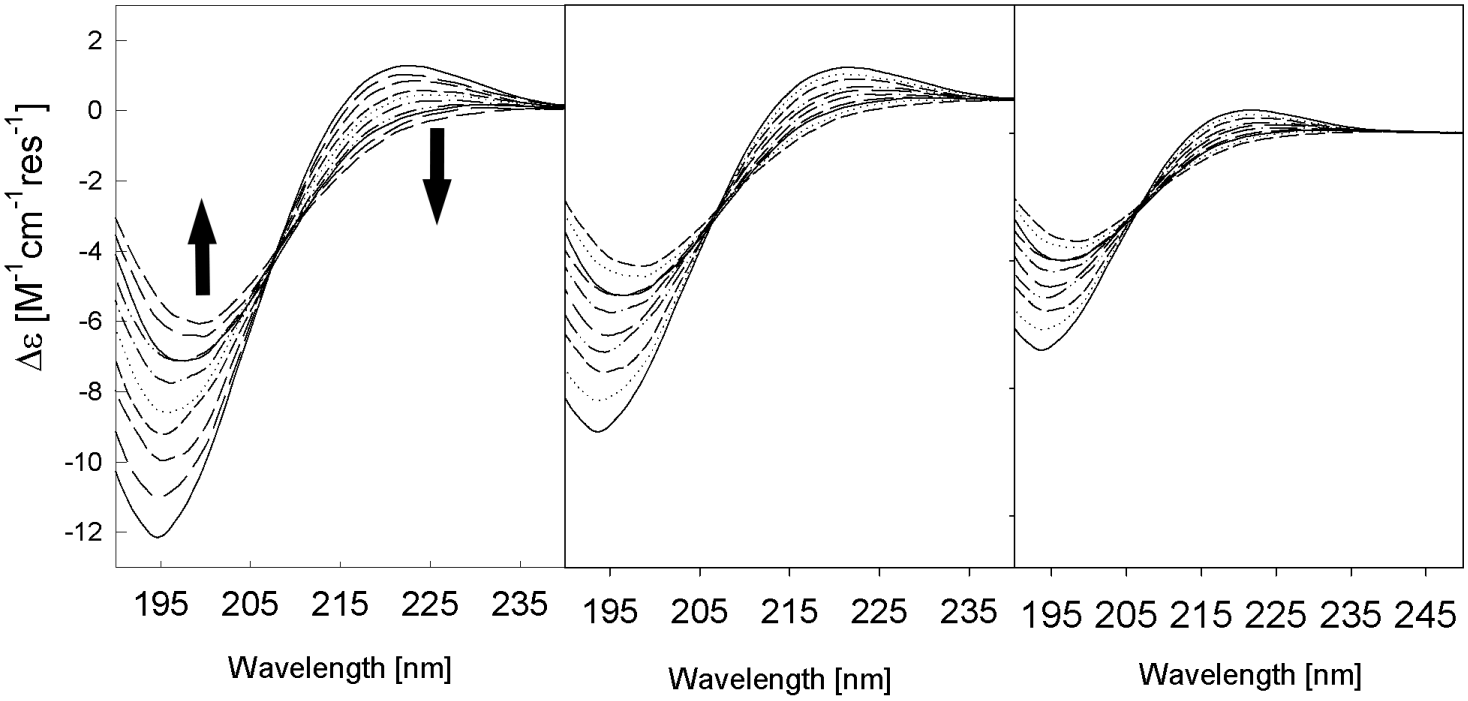


Figure III.2. A temperature dependent UV-CD spectra of peptides (a) $(GPP)_4$ (b) $(GPA)_4$ and (c) $(GPA)_2G(GPA)_2$ in 85/15 methonol/acetic acid solvent is shown. Arrows indicate increasing temperature from 5° to 95° C.

Protein Structure Prediction and Homology Modeling

To investigate how the glycine insertion found in the TrpA sequence would affect the secondary structure of the peptide, the 13 amino acid peptide sequence $(GPA)_2G(GPA)_2$ and the 12 amino acid peptide sequence $(GPA)_4$ with a mutation of the TrpA protein without the glycine insertion were used to predict the structure (Figure III.3).

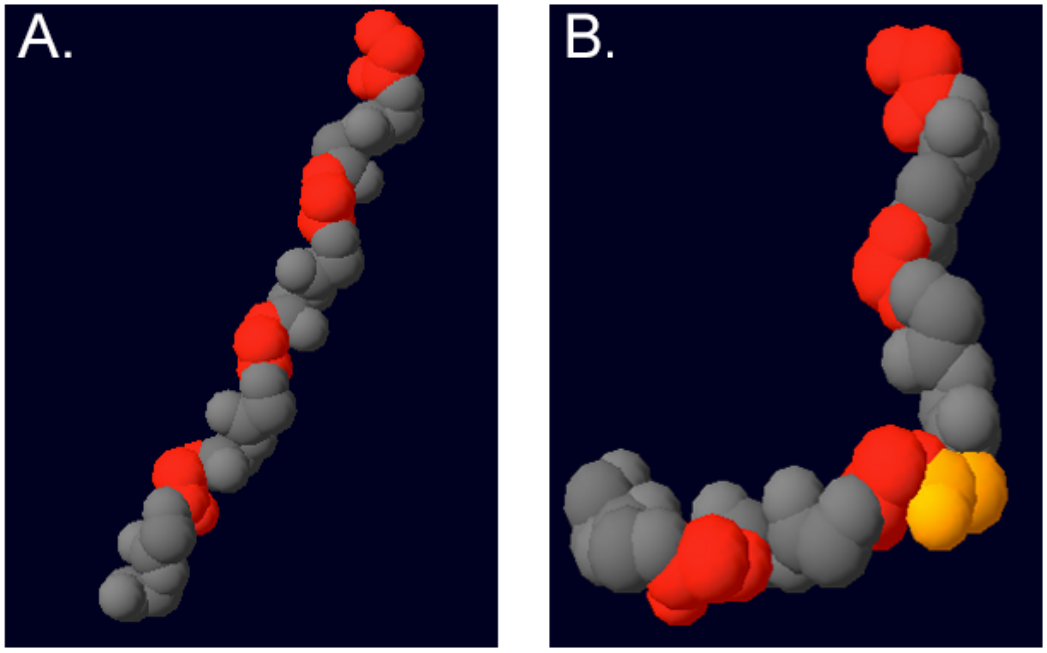


Figure III.3. Three dimensional protein prediction from I-Tasser of (a) $(GPA)_4$ and (b) $(GPA)_2G(GPA)_2$ viewed in DeepView. Solid, three dimensional Van der Waals representations of the backbone are shown. Glycine residues are highlighted in red. The additional glycine found between consecutive Gly-X-Y is highlighted in orange

The secondary structure predicted for both sequences were coils and aligned to the same collagen protein (PDM code 2klwA). The C-score, an estimate of the quality of the predicted model, was -0.80 for $(GPA)_2G(GPA)_2$ and -0.62 for $(GPA)_4$ confirming that models are correctly folded (>-1.5) (Roy 2010). The TM-score for both proteins also indicated that they are useful structural analogs (>0.5) that can be used to determine the protein family the sequence is related to; $(GPA)_2G(GPA)_2$ has a TM-score of 0.61±0.14 and $(GPA)_4$ has a TM-score of 0.63±0.13 (Roy *et al.*, 2010). The same analysis was done with a longer, 31 amino acid long sequence (TrpA_31), from TrpA extended on either side of the original $(GPA)_2G(GPA)_2$ because the length of the collagen chain (PDM code 2klwA) identified as the best template for the 13 amino acid sequence is 30 amino acids in length (Figure III.4). The C-Score and TM-Score was -1.07 and 0.58±0.14, respectively.

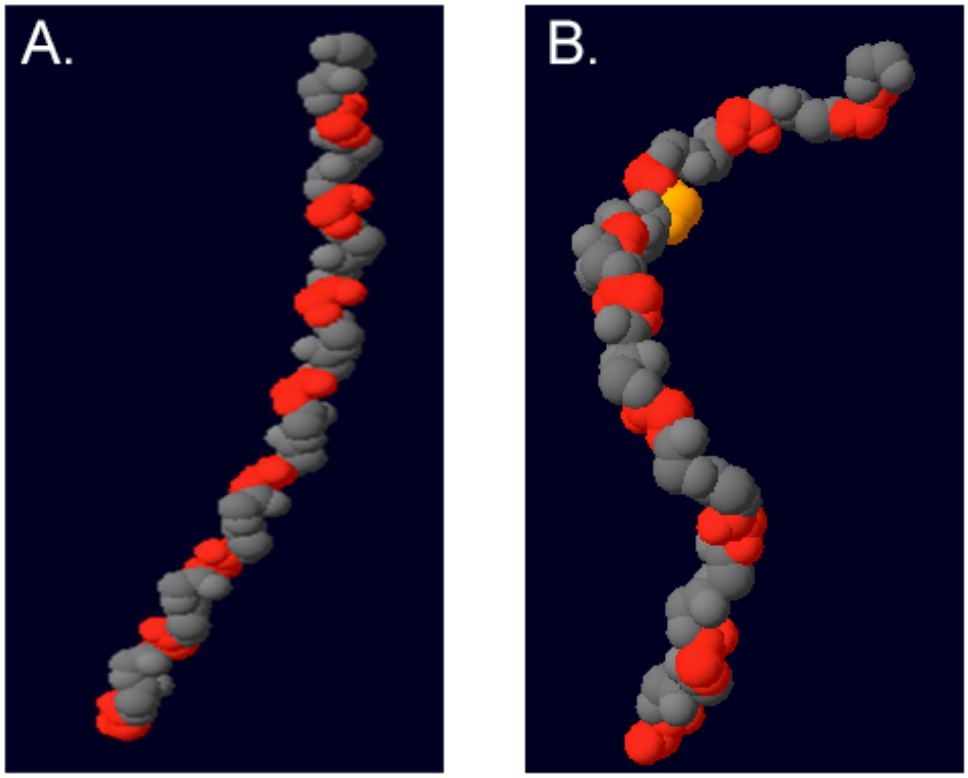


Figure III.4. Three dimensional protein prediction from I-Tasser of (a) PDB code 2klwA: $(GPK)_{10}$ and (b) TrpA_31: $(GPA)_6G(GPA)_4$ viewed in DeepView. Solid, three dimensional Van der Waals representations of the backbone are shown. Glycine residues are highlighted in red. The additional glycine found between consecutive Gly-X-Y is highlighted in orange

Using the TrpA_31 sequences the target sequence and collagen chain 2klwA as the template, a comparative protein model was constructed (Figure III.5). The identity between the two sequences was 64.28% verifying that it is an acceptable target-template for alignment and well above the 'twilight zone' (>30%) in which model quality becomes unreliable (Bordoli *et al.*, 2008, Schwede et al., 2003). The glycine insertion in the TrpA_31 target protein creates an obvious puckering or kink in the secondary structure

compared to the template structure (Figure 5b). The length between the first glycine (Residue 16) before the glycine insertion in TrpA_31 and the corresponding glycine in the 2kwlA (Residue 16) to next consecutive glycine (TrpA_31, Residue 20; 2kwlA, Residue 19) also displays an increase in length of 1.66 Å (TrpA_31: 10.99 Å, 2kwlA: 9.33 Å). The length of the protein models from the first glycine before the glycine insertion and the corresponding glycine in the template to the final glycine in both models (TrpA_31: Residue 16 to Residue 29, 2klwA: Residue 16 to Residue 28) exhibited increase in length of only 0.01 Å (TrpA_31: 20.13 Å, 2kwlA: 20.12 Å) (Figure 5b). The angle was measured between glycine residues 16 and 19 in 2klwA and residues 16 and 20 in TrpA. The angle was increased by 16.62° (TrpA: 141.8°, 2kwlA: 125.18°). The change in angle due to the glycine insertion distorts the helical structure in the target model (Fig 5c).

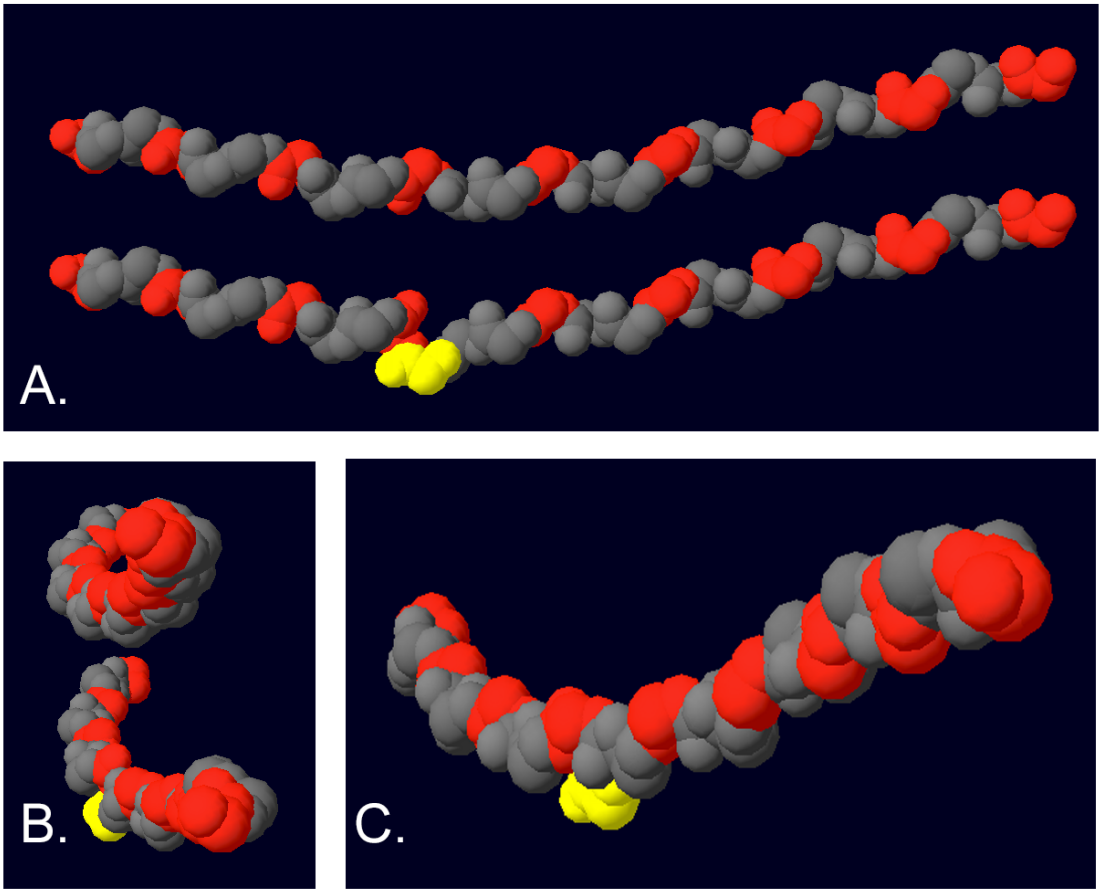
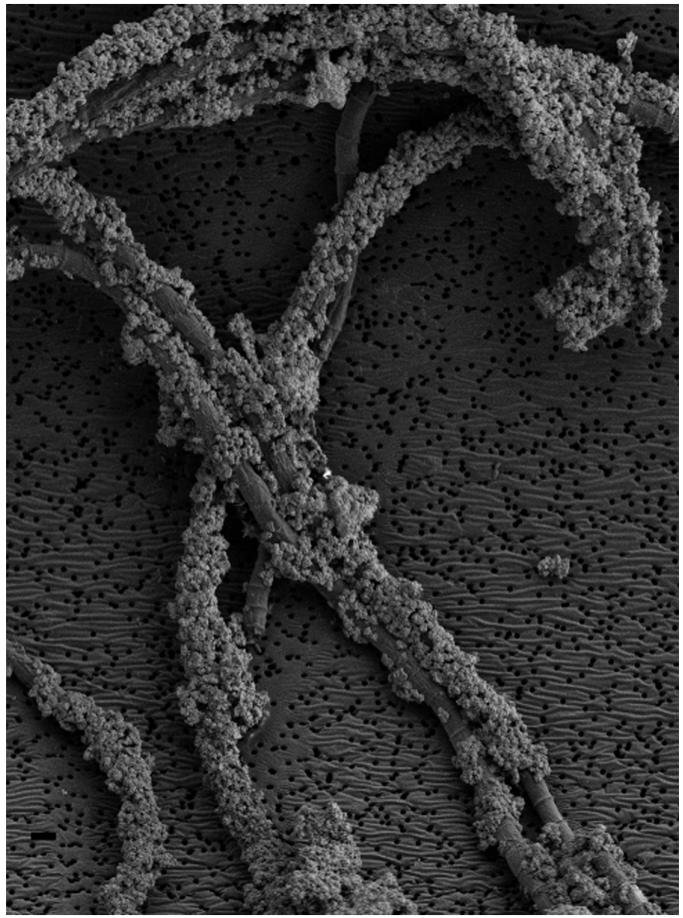


Figure III.5. Homology model of TrpA_31: $(GPA)_6G(GPA)_4$ using PDB code 2klwA: $(GPK)_{10}$ viewed in DeepView. Solid, three dimensional Van der Waals representations of the backbone are shown. Glycine residues are highlighted in red. The additional glycine found between consecutive Gly-X-Y is highlighted in yellow. 2klwA (top) and TrpA_31 (bottom) side by side (a) front view and (b) side view. Perturbation caused by the addition of glycine can be seen (c).

Silver enhancement of the secondary gold colloidal antibodies allowed for visualization of collagen localization on the surface of *T. erythraeum* cells (Fig 6). The distribution of the collagen protein on the surface of trichomes as indicated by silver enhancement indicates there is no specific pattern of localization. Controls for silver enhancement were negative for labeling. Possible collagen fiber like structures ~20nm in diameter can be seen on the surface (Fig 7).

A.



B.

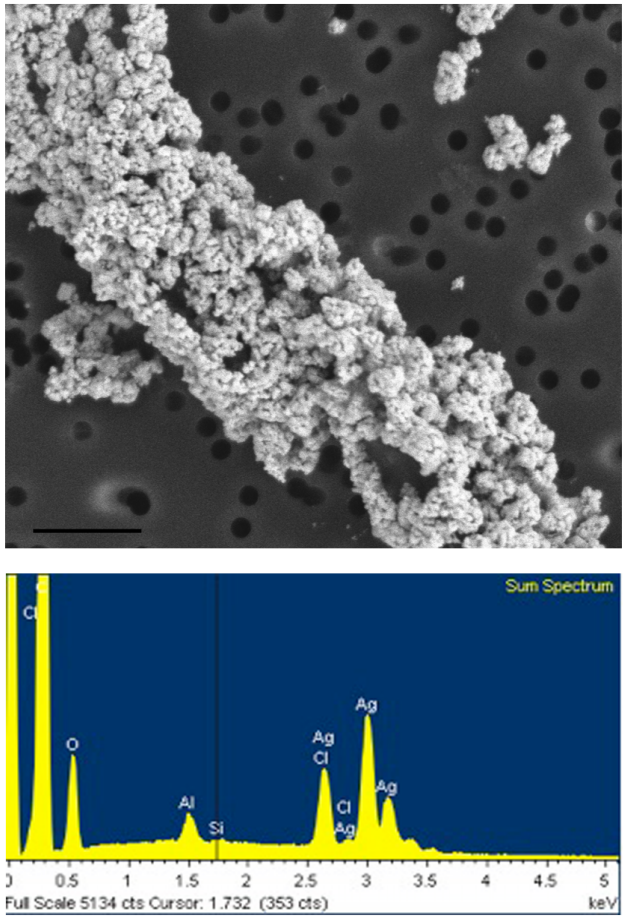


Figure III.6. Carbon coated silver enhancement treated trichomes that were immunogold labeled and probed for collagen (a) 500x (a) 3000x (c) X-ray confirmation of silver labeling. Scale bar is 5um.

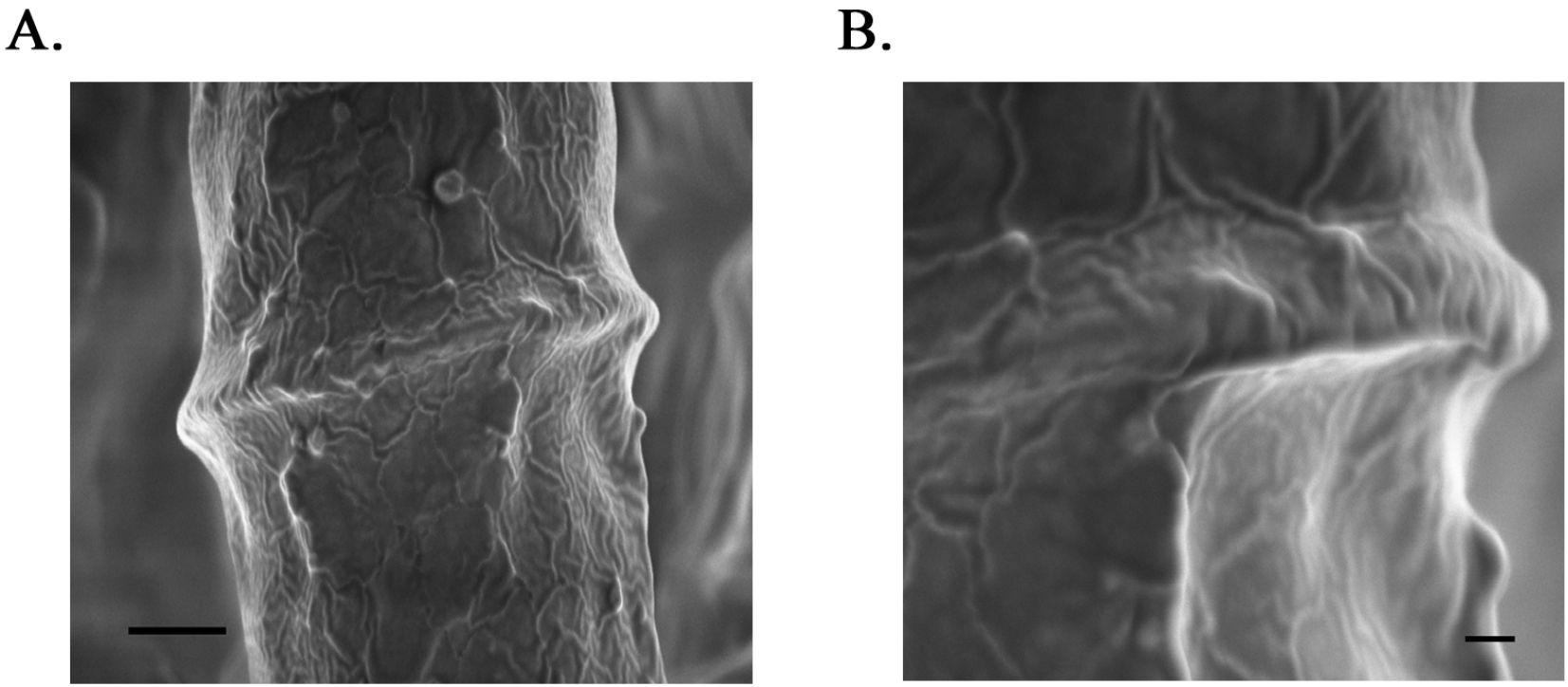


Figure III.7. SEM of carbon coated *T. erythraeum* trichomes without silver enhancement, (a) 10,000x, scale bar = 1µm (b) 30,000x, scale bar = 200nm

III.IV Discussion

Collagens have evolved to provide a variety of structures such as fibrils, microfibrils and meshwork (Kielty & Grant, 2003). Classification of collagen into subfamilies, fibril-forming or non fibril-forming, takes into account a number of factors including the primary sequence, the length of the collagen chains and the supramolecular quaternary structures that form. We have combined these factors to make a prediction about the structural classification of TrpA collagen.

A number of collagen proteins contain short collagenous domains that can be either continuous or interrupted that form structures including fibril-associated collagens with interrupted triple helical domains (FACIT), transmembrane collagens and multiplexin collagens (Kielty & Grant, 2003). The triglycine repeat length of the TrpA protein found in *T. erythraeum* is approximately 10% longer than that from any other sequenced collagen from eukaryotic or prokaryotic organisms (Layton et al., 2008). Based on length alone, TrpA can be excluded from creating a structure that is typically formed by short collagen domains.

Based on whole genome sequencing *T. erythraeum* contains only one gene encoding a collagen-like protein predicting that only a homotrimeric helix would form. Previous work from this laboratory confirms that there is only one protein in this organism that cross reacts with collagen Type I-IV antibody (Price and Anandan). Using amino acid sequences of the 80 vertebrate, invertebrate and prokaryotic collagens, we have gained some insight into what structure this collagen-like protein may form based on its grouping with other collagen proteins. A maximum likelihood approach was used

because it provides the highest aggregate probability, which most likely reflects the true phylogenetic tree (Krane & Raymer). *T. erythraeum* collagen grouped with the clade that included nonfibrillar collagens (Figure 1). Interestingly, all of the invertebrate fibrillar collagens also grouped amongst the nonfibrillar collagens.

The reason the invertebrate fibrillar collagens grouped with the nonfibrillar collagens is due to the fact that invertebrate fibrillar collagen sequences contain imperfections in their sequence. The presence of an additional glycine is characteristic of invertebrates and lower vertebrate chains. For example, Gly-X-Gly of *Riftia* or Gly-X-Y-Z seen in sponges have been shown not to affect molecular stability and fibrillar integrity in these proteins (Sicot *et al.*, 1997, Exposito & Garrone, 1990). In *Riftia* α chain, there is a single replacement of a glycine to an alanine in the regular Gly-X-Y repeat that should cause a disturbance in the triple helical rigidity, however it does not prevent triple helical folding although it does decrease thermal stability (Mann *et al.*, 1992). The presence of imperfections in invertebrate collagens does not impair its fibril function (Mann et al., 1992).

The low temperature at which these organisms that contain the imperfect fibril forming collagen sequences are found has been implicated as a factor in preventing these imperfections from affecting the structure (Exposito et al., 2002). Considering *T. erythraeum* is found in similar temperature environments, it is possible that although phylogenetically the collagen-like protein is in a clade of nonfibrillar collagens, the imperfection found in the sequence does not impair fibril formation similar to the fibril forming invertebrate collagens it is groups with in this clade (Figure III.1).

To determine if the glycine imperfection in the *T. erythraeum* collagen sequence would prevent triple helical conformation, synthetic peptides were synthesized based on the TrpA sequence. A $(GPP)_4$ peptide sequence was used to represent a sequence more similar to that of vertebrate collagen, however without the hydroxyproline in the Y position. In vertebrates, the presence of hydroxyproline is a stabilizing force in triple helices. However, upon the identification of collagen-like sequences in prokaryotes it has been shown that hydroxylation of the proline residue is not a requirement for triple helical structure (Mohs *et al.*, 2007). Collagen-like proteins in *Bacillus anthracis* and *Streptococcus pyogenes* have been shown to self-assemble and form triple helical structures in the absence of hydroxyproline (Mohs et al., 2007). $(GPA)_2G(GPA)_2$ is representative of the collagen sequence from residue 723 to residue 735 of the TrpA protein, while the $(GPA)_4$ is a mutation of the *T. erythraeum* sequence without the imperfection. CD spectra reveals that $(GPA)_2G(GPA)_2$ has a lower PPII content and is less stable than both $(GPP)_4$ and $(GPA)_4$ in which there is no glycine insertion.

The circular dichroism results corresponds with the predicted model of TrpA from I-TASSER and the homology model of TrpA from Swiss-Model which both indicate that the additional glycine would cause a disruption to the secondary structure of the TrpA collagen chain decreasing its stability. However, $(GPA)_2G(GPA)_2$ still exhibit PPII content and thus show no indication that it would be incapable of self-assembling to form collagen triple helices. The CD spectra indicates that at lower temperatures, the PPII content and stability of the $(GPA)_2G(GPA)_2$ would increase which aligns with previous research showing that collagen proteins containing interruptions found in invertebrates

growing at lower temperatures form stable, fibril forming collagens (Exposito et al., 2002).

Previous work from this laboratory has shown that the TrpA protein is localized to the junctions between cells along the multicellular trichome of *T. erythraeum* (Price and Anandan). In an effort to resolve the function of the collagen protein Scanning Electron Microscopy was used to observe the structure that the collagen protein was forming on the cells. In accordance with previous research done by Layton et.al, TrpA was labeled on the surface of the trichomes (Layton et al., 2008). Silver enhancement with SEM revealed that the collagen protein on the surface of the cells is not localized to a specific region of the trichome surface. This pattern of localization indicates that collagen in *T. erythraeum* may also function in a sheath formation on the trichome surface indicating a secondary function for TrpA (Figure 3). The presence of a collagen protein may function in cell-cell adhesion of trichomes to one another in colony formation. It is also possible that heterotrophic bacteria utilize the presence of this protein on the surface for adherence.

Though the glycine insertion introduces instability into the structure, this research confirms that the sequence found in the *T. erythraeum* genome will create a collagen molecule. While more research needs to be done, identifying the ability of the sequence to form collagen molecules introduces the possibility of exploiting this collagen protein for use in biopolymer engineering.

Chapter IV. Towards creating a TrpA knockout

IV.I Introduction

Cyanobacteria are one of the most diverse and vastly distributed phyla of bacteria (Shih *et al.*, 2013). Their unique characteristics make them models for studying photosynthesis, genetic control of gene expression, photoregulation, cell differentiation, primary production and molecular evolution (Koksharova & Wolk, 2002). Genetic manipulation of these organisms is imperative for both basic research purposes and for biotechnological applications (Koksharova & Wolk, 2002). Currently, there are 35 completely sequenced genomes of cyanobacteria available for analysis representing merely a fraction of all cyanobacteria (Cyanobase). Of those, few have been successfully genetically manipulated including *Synechococcus* PCC 7942, *Anabaena*PCC 7120, *Nostoc punctiforme* and *Fremyella diplosiphon* (Golden *et al.*, 1987, Wolk *et al.*, 1984, Flores & Wolk, 1985, Chiang *et al.*, 1992).

The transfer and uptake of genetic information in the form of DNA by bacterial cells is the crux of genetic manipulation, which can be achieved in bacterial systems through transformation, conjugation or transduction (Johnsborg *et al.*, 2007). DNA transfer in cyanobacteria has been achieved through natural transformation, artificial transformation (electroporation) and conjugation (Elhai, 1994). Of the cyanobacteria that have been successfully manipulated, many unicellular cyanobacteria are known to be naturally transformable by exogenous DNA (Koksharova & Wolk, 2002). *Synechococcus elongatus* PCC7942 has become a model for research on photosynthesis and circadian rhythms as a result of its natural transformative ability (Golden & Sherman, 1984, Kondo

et al., 1993, Anandan & Uram, 2004). However, natural transformation has yet to be demonstrated in filamentous cyanobacteria (Flores *et al.*, 2008).

Conjugation has been shown to be an effective DNA transfer system for the heterocystous, filamentous cyanobacteria *Anabaena* PCC7120 and *Nostoc punctiforme* (Wolk et al., 1984, Flores & Wolk, 1985). Interspecific conjugation requires that the DNA be in plasmid form thus electroporation was found to be a simpler alternative because it allows transfer of various forms of DNA (Elhai, 1994). Restriction modification systems in cyanobacteria are the main barrier to genetic manipulation (Zhao *et al.*, 2006). Regardless of the way the DNA is transferred into cyanobacterial cells, degradation of transferred DNA by restriction endonucleases native to the recipient strain will prevent effective DNA uptake (Koksharova & Wolk, 2002).

Of the six *Trichodesmium* spp., *Trichodesmium erythraeum* is the only species to have been completely sequenced. *T. erythraeum* a filamentous, nonheterocystous, marine cyanobacterium that is notorious for being difficult to maintain in laboratory cultures and requiring special handling for successful cultivation (Ohki & Fujita, 1982, Borstad & Borstad, 1984, Prufert-Bebout *et al.*, 1993, Chen *et al.*, 1996). Cultivation of this organism in the laboratory has allowed for experiments to be performed investigating the physiology and molecular biology of *T. erythraeum* (Chen et al., 1996). Credited with a significant amount of marine nitrogen fixation, significant research has been focused on the ecological impact of *T. erythraeum* as a primary producer (Capone et al., 1997). Despite successful cultivation, a genetic system has yet to be developed in *T. erythraeum* limiting the ability to study the impact of genetic manipulation within the *T. erythraeum* cells, and limiting the ability to create gene knockouts and recombinant strains.

This study aimed at developing a genetic system in *T. erythraeum* to allow for genetic manipulation ultimately targeting the expression and regulation of TrpA. As a first step to developing a genetic system in this organism, we have successfully determined which antibiotics can be used as selective markers for screening of successful DNA transfer. Using both bioinformatic analysis of the sequenced genome and empirical data, we have also determined that *T. erythraeum* is most probably able to undergo natural transformation.

IV.II Materials and Methods

Bioinformatic analysis

Amino acid sequences of competence related proteins ComEA, ComEC, ComA, ComQ, ComGC, ComFA, PilD, PilF, and PilG from the naturally transformable bacterial species *B. subtilis, S. pneumoniae, N. gonorrhoeae,* and *H. influenzae* (Claverys & Martin, 2003) were acquired from the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov) and individually compared against the *S. elongatus PCC 7942* and *T. erythraeum IMS101* genome using BLASTP.

Bacterial strains and plasmid preparation

T. erythraeum cultures were grown as described in Chapter 2. Z-competent™ *E. coli* DH5α (Zymo Research, Irvine, CA, USA) were transformed with pUC19 (Invitrogen) and grown according to manufacturers protocol. The plasmid was purified using QIAprep Spin Miniprep Kit (Qiagen) according to manufacturers protocol.

Reverse Transcription and PCR Amplification

RNA extraction and cDNA synthesis were done as described in chapter 2. Using genomic sequence data, Primers were synthesized for *comEA*, *comEC/Rec2*, *dprA* and *tery_4117* (Table IV.1). Following amplification, the RT-PCR products were subjected to 1.75% gel electrophoresis with ethidium bromide staining to visualize the DNA.

Table 1. Primers used for RT-PCR, 5'→ 3'

ComEAF	CCATTGCCACAAGAACCTCT
ComEAR	TGAACCGCCACATACACAGT
ComECF	ATCGCTCTGCTGGCAAGTAT
ComECR	ACCAAAGTTGGGCTTTTCCT
DprAF	ATCCCAATTTTTGGACACCA
DprAR	ACGACGACCATATTCCGAAG
Tery4117F	CGATGCTGATGGTTCTGATG
Tery4117R	CATCCTCCTTTGCCGTAGAA

Minimum Inhibitory Concentration (MIC) Assays

Ampicillin, kanamycin, streptomycin, spectinomycin and chloramphenicol (Sigma, xx, USA) stocks were made resuspended in RMP media. 200 µl *T. erythraeum* culture were aliquoted into polystyrene 96 well plates (USA Scientific) and incubated with 0-10 µg/ml of each antibiotic in increments of 2 µg/ml. Direct microscopic observations were made every 24 hours over a course of 5 days using a LeicaEZ4 dissecting microscope (Leica USA). Wells were scored based on the presence of healthy or unhealthy trichomes or the

absence of trichomes, as well as any phenotypic changes. These assays were conducted in duplicate with at least three independent trials.

Transformation of T. erythraeum

From cultures 19-21 days post inoculation, 2ml of cells were removed and centrifuged at 3000 x g for 15 minutes. The cells were resuspended in ice-cold RMP media followed by centrifugation at 3000 x g for 15 minutes. The cells were resuspended in 200 µl ice-cold RMP media and incubated overnight with 150 ng of pUC19 (IMS101+P) under normal culture conditions. After 24 hours of incubation, cells were aliquoted into polystyrene 96-well plates (USA Scientific) and inoculated with ampicillin to a final concentration of 2 µg/ml. The cells were observed every 24 hours for 5 days as described above. As a control for antibiotic treatment, cells were prepared for DNA uptake in same manner except without the addition of pUC19 (IMS101-P) and fresh *T. erythraeum* cells were also inoculated with ampicillin. To control for the effect of preparing cells for DNA uptake, cells both with and without pUC19 were incubated in RMP media without ampicillin. All wells including controls were scored as described for MIC assays. These assays were conducted in duplicate with at least three independent trials.

IV.III Results

Identification of Competent Genes in T. erythraeum IMS101 Genome and their Expression

ComEA, ComEC, ComA, ComQ, ComGC, ComFA, PilD, PilF, and PilG proteins have been shown to be required for competence in model transformative bacterial cells,

including *B. subtilis*, *E. coli*, *N. gonorrhoeae*, and *S.pneumoniae* (Claverys & Martin, 2003). DprA is required for the protection of internalized ssDNA from nucleases (Chen et al., 1996). BLASTP searches revealed that almost all of these proteins are encoded in the *Synechococcus elongatus* PCC 7942 and *T. erythraeum IMS101* genomes (Table IV.2). *S. elongatus* PCC 7942 was included in the quest for competence genes because it is a well characterized naturally transformable cyanobacterium, and serves as a positive control for these analyses. Tery_0225 and Tery_4491 in *T. erythraeum* are homologous to signaling proteins that are involved in regulating timing of the development of competence in in *S. pneumoniae* and *B. subtilis.* Tery_0225 has 32% identity to ComA export protein in *S. pneumoniae* and 28% identity to ComA sensing protein in *B. subtilis.* DprA in *T. erythraeum* has a 43% identity to the DNA protection protein in *S. pneumoniae,* 38% to this protein in *H. influenzae* and *B. subtilis* and 35% to Dpr in *N. gonorrhoeae.*

All of the DNA uptake machinery proteins except the DNA translocase protein identified in the model transformative bacteria that exhibit competence had homologues in *T. erythraeum*. Tery_4117 had a 45-48% identity to the peptidoglycan crossing protein in *N. gonorrhoeae* and H. influenza. ComEA had a 42-45% identity with the DNA binding protein in all of the model bacteria. ComEC/Rec2 had a 22% identity with the transmembrane channel protein in *B. subtilis*, 21% with *S. pneumoniae*, and 26% with *H. influenzae*. Tery_2683, Tery_1748, and Tery_1745 were homologous to the prepilin assembly proteins in almost all of the model competent bacteria. The prepilin peptidase protein, Tery_2683, had a 26%-35% identity *B. subtilis, N. gonorrhoeae,* and *H. influenzae.* Tery_1748 had a 29%-42% identity to the traffic NTPase proteins in *B. subtilis, N. gonorrhoeae, S. pneumoniae* and *H. influenzae.* The polytopic membrane

protein found in *B. subtilis, N. gonorrhoeae,* and *H. influenzae* has a percent identity of 18%, 35% and 25% respectively to Tery_1745. *T. erythraeum* competence proteins have the highest percent identity with *S. elongatus* PCC 7942 proteins involved in signaling function, DNA protection and pore assembly. From these data, we believe that *T. erythraeum* has the potential to be a naturally transformable bacterium capable of DNA uptake.

Table IV.2. Percent homology of competent proteins in *T. erythraeum* with known model competent bacteria.
(*) No clear orthologue in model transformable gram-negatives

	T. erythraeum	*B. subtilis*	*S. pneumoniae*	*N. gonorrhoeae*	*H. influenzae*	*S. elongatus PCC7942*
Signaling function						
Export	Tery_0225	ComQ	ComA (32%)	---	---	Synpcc7942_1905 (39%)
Sensing	Tery_4491	ComA (28%)	ComD-E	---	---	Synpcc7942_1860 (67%)
DNA protection						
	DprA	Smf (38%)	DprA (43%)	DprA (35%)	DprA (38%)	DprA (46%)
DNA Uptake Machinery						
Peptidoglycan crossing	Tery_4117	ComGC	ComGC	PilE (45%)	Hi0299 (48%)	Pila (38%)
DNA binding	ComeA	ComEA (45%)	ComEA (48%)	ComE(45%)	Hi1008 (42%)	Synpcc7942_0869 (36%)
Transmembrane channel	ComEC/Rec2	ComEC (22%)	ComEC (21%)	ComA	Rec2 (26%)	ComEC/rec2 (34%)
DNA translocase	*	ComFA	ComFA	*	*	*
Pore Assembly						
Prepilin peptidase	Tery_2683	ComC (35%)	Ccl	PilD (34%)	Hi0296 (26%)	Synpcc7942_1935 (52%)
Traffic NTPase	Tery_1748	ComgA (38%)	ComgA (29%)	PilF (42%)	PilB(36%)	Synpcc7942_2071 (55%)
Polytopic membrane	Tery_1745	ComgB (18%)	ComgB	PilG (35%)	CglB (25%)	Synpcc7942_2069 (52%)

To confirm that the competent genes identified in the *T. erythraeum* genome were expressed, the expression of select genes, *comEA*, *comEC/Rec2*, *DprA* and *Tery_4117*, was investigated using RT-PCR on Day 21 cultures. RT-PCR confirms that all of these genes selected are transcribed (Fig IV.1). A putative model of how the *T. erythraeum* competence proteins may interact during DNA uptake is given in Figure 2, based on previous models that represent the probable arrangement of the proteins involved in competence (Fig IV.2).

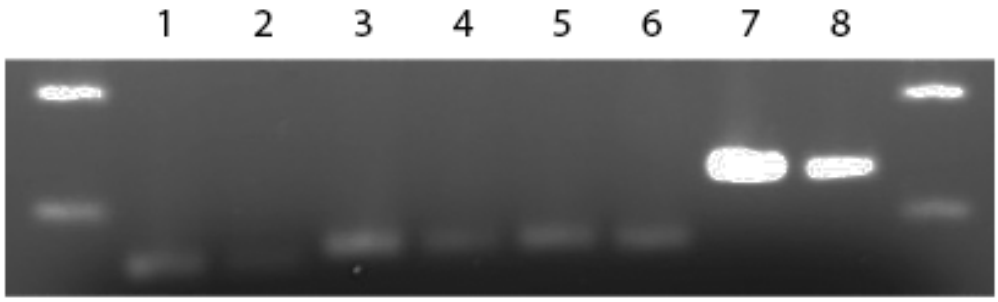


Figure IV.1. Expression of *T. ertythraeum* competence genes. PCR product of genomic (Lanes 1, 3, 5, and 7) and cDNA (Lanes 2, 4, 6, and 8). Primers for *comEA* (Lanes 1, 2), *comEC/Rec2* (Lanes 3, 4), *dprA* (Lanes 5, 6) and *Tery_4117* (Lanes 7, 8).

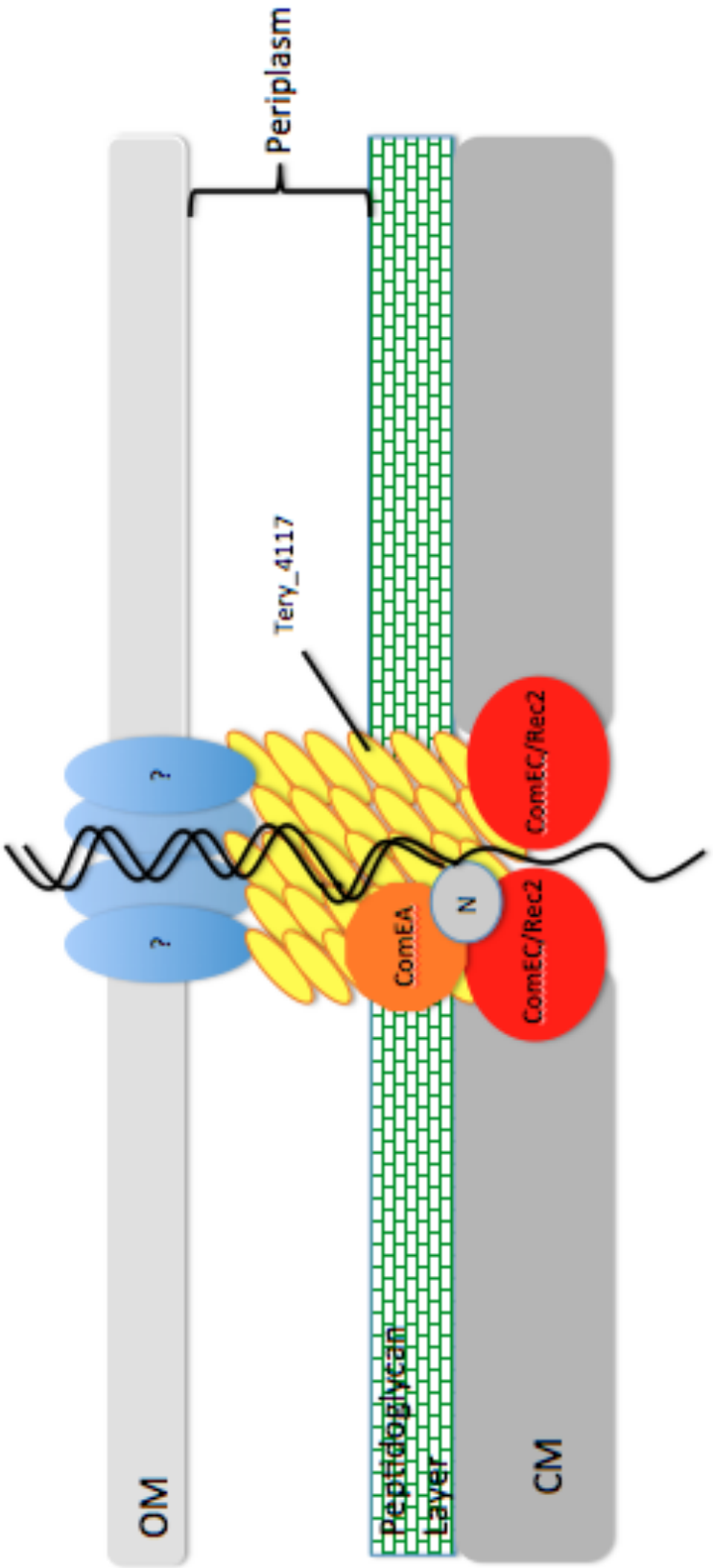


Figure IV.2. Model of DNA Uptake Model with *T. erythraeum* competence genes. Based on Clavery 2003 proposed model.

Identification of Antibiotic Susceptibility in in T. erythraeum IMS101

As a first step to developing a genetic system, we investigated suitable antibiotic markers for use in this organism. Trichomes incubated with antibiotics were scored as healthy, unhealthy or dead based on their appearance. Wells containing healthy cells had long trichomes, wells described as unhealthy contained only fragments of trichomes and wells scored as dead had no trichomes present. Representative pictures of healthy, unhealthy and dead trichomes are shown in figure IV.3.

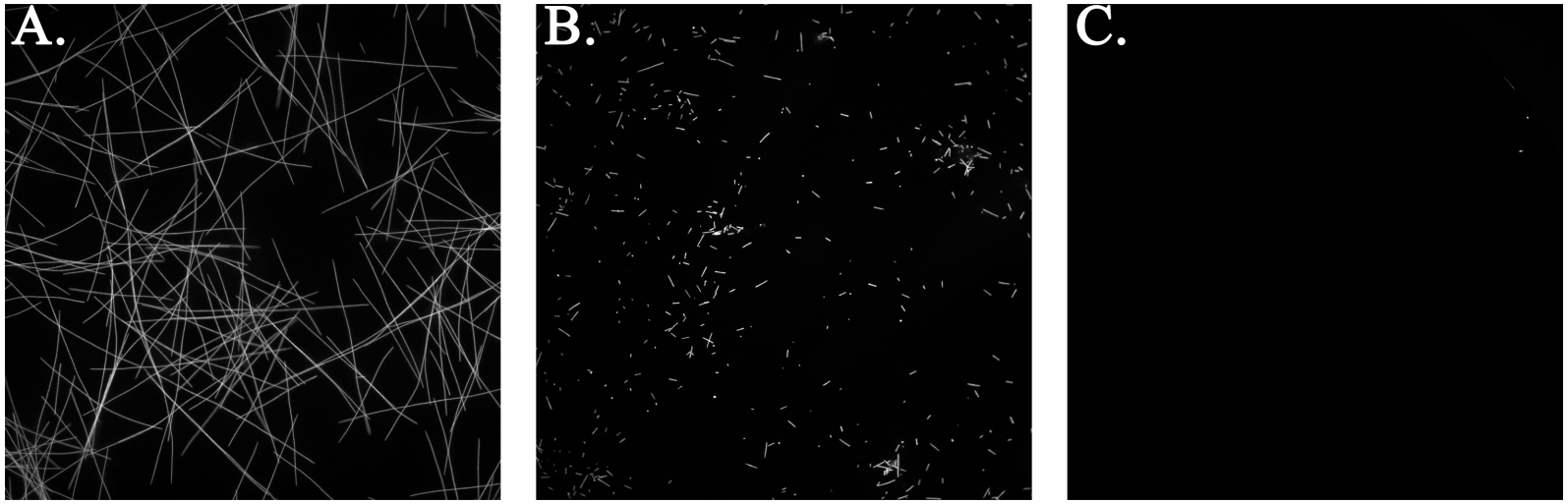


Figure IV.3. Representative trichome presence or absence for scoring MIC and Transformation Assay. (a) healthy trichomes (b) unhealthy trichomes (c) dead.

Trichomes in the presence of ampicillin began to fragment on day three at every concentration (2μg/ml to 10μg/ml) used and were dead by day four (Table 3,b) indicating that *T. erythraeum* is extremely susceptible to ampicillin. Trichomes incubated with streptomycin began to exhibit degradation of trichomes at 4.0μg/ml on day four and were dead on day five at antibiotic concentrations of 6μg/ml to 10μg/ml (Table 3,d). Incubation with the antibiotics spectinomycin and chloramphenicol affected the trichomes in similar ways; by day two the trichomes were unhealthy at every concentration and by day four there were no trichomes present (Table IV.3a,e,f). In contrast, the trichomes were completely unchanged over the course of the study in the presence of kanamycin (Table IV.3a,e), and appeared healthy, indicating that at the concentrations used for this study, *T. erythraeum* is not susceptible to kanamycin treatment. From these data, ampicillin was chosen as a suitable antibiotic marker, since *T. erythraeum* cells are extremely susceptible to this antibiotic even at the very low concentrations used in the study.

Table IV.3. MIC Assays. (++) indicates presence of healthy trichomes, (+) indicates presence of unhealthy trichomes (-) indicates absence of trichomes from well

	Day 1	Day 2	Day 3	Day 4	Day 5
Ampicillin	++	++	+	-	-
Kanamycin	++	++	++	++	++
Streptomycin	++	++	++	+ ≥ 4.0μg/ml	-
Spectinomycin	++	+	+	-	-
Chloramphenicol	++	+	+	-	-

Ampicillin	Day 1	Day 2	Day 3	Day 4	Day 5
0.0 μg/ml	++	++	++	++	++
2.0 μg/ml	++	++	+	-	-
4.0 μg/ml	++	++	+	-	-
6.0 μg/ml	++	++	+	-	-
8.0 μg/ml	++	++	+	-	-
10.0 μg/ml	++	++	+	-	-

Kanamycin	Day 1	Day 2	Day 3	Day 4	Day 5
0.0 μg/ml	++	++	++	++	++
2.0 μg/ml	++	++	++	++	++
4.0 μg/ml	++	++	++	++	++
6.0 μg/ml	++	++	++	++	++
8.0 μg/ml	++	++	++	++	++
10.0 μg/ml	++	++	++	++	++

Streptomycin	Day 1	Day 2	Day 3	Day 4	Day 5
0.0 μg/ml	++	++	++	++	++
2.0 μg/ml	++	++	++	++	++
4.0 μg/ml	++	++	++	+	+
6.0 μg/ml	++	++	++	+	-
8.0 μg/ml	++	++	++	+	-
10.0 μg/ml	++	++	++	+	-

Spectinomycin	Day 1	Day 2	Day 3	Day 4	Day 5
0.0 μg/ml	++	++	++	++	++
2.0 μg/ml	++	+	+	-	-
4.0 μg/ml	++	+	+	-	-
6.0 μg/ml	++	+	+	-	-
8.0 μg/ml	++	+	+	-	-
10.0 μg/ml	++	+	+	-	-

Chloramphenicol	Day 1	Day 2	Day 3	Day 4	Day 5
0.0 μg/ml	++	++	++	++	++
2.0 μg/ml	++	+	+	-	-
4.0 μg/ml	++	+	+	-	-
6.0 μg/ml	++	+	+	-	-
8.0 μg/ml	++	+	+	-	-
10.0 μg/ml	++	+	+	-	-

Transformation of T. erythraeum IMS101 with pUC19

Based on the above results, we next chose to incubate *T. erythraeum* with the antibiotic ampicillin, in the presence of a plasmid carrying an ampicillin resistance gene to test if the trichomes would show uptake of the plasmid DNA and thereby acquire ampicillin resistance**.** For these experiments, pUC19 was chosen based on the antibiotic resistant cassette it contains, ampicillin. Trichomes incubated with and without pUC19, IMS101+P and IMS101-P respectively, in RMP media without ampicillin were healthy throughout the duration of the assay. IMS101-P began to exhibit an unhealthy phenotype by day 2 and IMS101+P by day 3 in the presence of the antibiotic. By day four, IMS101-P trichomes were dead in the presence of ampicillin but trichomes remained healthy in untreated RMP media. IMS101+P trichomes were still present in the presence of ampicillin through the duration of the assay. Compared with IMS101+P trichomes in RMP, IMS101+P in the presence of RMP and ampicillin were fragmented (Table IV.4) indicating that come of the cells along the trichome were susceptible to ampicillin.

Table IV.4. Competency Assay Results. (++) indicates presence of healthy trichomes, (+) indicates presence of unhealthy trichomes (-) indicates absence of trichomes from well

Day 1	-Amp	+Amp
IMS101-P	++	++
IMS101+P	++	++

Day 2	-Amp	+Amp
IMS101-P	++	+
IMS101+P	++	++

Day 3	-Amp	+Amp
IMS101-P	++	+
IMS101+P	++	+

Day 4	-Amp	+Amp
IMS101-P	++	-
IMS101+P	++	+

Day 5	-Amp	+Amp
IMS101-P	++	-
IMS101+P	++	+

IV.IV Discussion

The presence of genes in a bacterial genome that are the result of lateral or horizontal gene transfer is an indicator of an organisms ability to take up and incorporate exogenous DNA (Chen *et al.*, 2005). The process of natural transformation has been identified in very few bacteria, likely due to the inability to pinpoint the specific

physiological conditions that signal the need for an organism to be competent (Chen & Dubnau, 2004). Once those conditions for transformation have been met in competent gram-negative bacteria, incoming DNA must cross the outer membrane, cell wall and cytoplasmic membrane before reaching the cytoplasm and nucleoid region (Chen & Dubnau, 2004). Natural competence in bacteria has been indicated to be beneficial due to increasing genetic diversity but also in providing DNA as a nutritional resource (Finkel & Kolter, 2001). Using BLASTP and bioinformatic analyses, we have identified and demonstrated that the majority of proteins involved in common bacterial competence systems are encoded in the *T.erythraeum* IMS101 genome (Table 1).

Competence in *T.erythraeum* could be signaled by the cell density responsive quorum sensing protein, Tery_4491. Tery_4117 prepilin proteins are likely processed by Tery_2683 prepilin peptidase and then translocated and assembled by Tery_1748, a traffic NTPase and Tery_1745, a polytopic membrane protein. The processed pilin proteins could create a pilin-like complex that would allow for DNA to transverse the periplasm and peptidoglycan layer (Johnsborg et al., 2007, Claverys & Martin, 2003). It has been shown that internalization of DNA depends on the presence of Type IV pili in naturally transformable cyanobacteria (Johnsborg et al., 2007). The DNA passing through the complex would bind the DNA binding receptor ComEA and direct the DNA to a nuclease that would degrade one strand and deliver the other strand through a channel formed by ComEC/Rec2. The DprA protein would then act to protect the incoming DNA from nucleases, which would be necessary because *T. erythraeum* has 15 Restriction Modification (RM) systems (Ives *et al.*, 1992, Ando *et al.*, 2000)(New England Biolabs).

Our RT-PCR analysis demonstrated expression of the *comEA*, *comEC/Rec2*, *dprA* and t*ery_4117* genes in cells from cultures 19-21 days post inoculation. Figure 1 illustrates the putative competence complex that could be formed in *T.erythraeum*. Identification of these competency proteins that show significant identity to proteins in previously well characterized systems suggests that *T.erythraeum* is able to incorporate DNA into its cells via natural transformation under the appropriate conditions.

Genetic manipulation requires that identifiable markers be used to confirm DNA uptake. Antibiotic assays showed that *T.erythraeum* is susceptible ampicillin, streptomycin, spectinomycin, and chloramphenicol. A gene that confers antibiotic resistance to any of the antibiotics *T.erythraeum* is susceptible to could be used as a marker to identify if DNA uptake has occurred. Interestingly, *T.erythraeum* was resistant to kanamycin at the concentration range in the antibiotic assays although other cyanobacteria are known to be susceptible to kanamycin (Andersen, 2005).

The DNA source for transformation assays, pUC19, was chosen because it is a widely used vector that contains an antibiotic resistance gene. It also has a prokaryotic origin of replication (ORI) to allow expression of the ampicillin resistant gene without incorporation into the chromosomal genome. This ORI is compatible with *Trichodesmium* as is the promoter for expression of resistance. DNA uptake was confirmed by the survival of IMS101+P cells in RMP with ampicillin on day 5 of the transformation assay while this was not the case for IMS101-P cells. Fragmentation of the IMS101+P cells in the presence of ampicillin could be the result of low transformation efficiency under the conditions used. *T. erythraeum* is filamentous, with individual cells along the trichome, thus a low transformation efficiency would result in

only a percentage of cells in the trichome showing transformation with the plasmid DNA. The cells along the trichome that was not successfully transformed would be susceptible to the ampicillin in the media and fragmentation of the trichome would result. It has been shown that relatively high concentrations of DNA are required for a large number of transformants in filamentous cyanobacteria (Thiel & Poo, 1989).

This study is the first to identify genes encoding proteins involved in natural competency in the *T. erythraeum* genome. Additionally, we have shown that DNA uptake is exhibited by *T. erythraeum*. Demonstration of natural transformation in *T. erythraeum* allows for the continued development of a genetic system in this organism that will allow the creation of recombinant strains, particularly gene knockout strains. This provides an invaluable tool for genetic manipulation to allow for creation a TrpA knock-out. Based on previous research TrpA is involved in cell to cell adhesion. (See chapters II & III) Creation of a knock-out strain would determine if this protein is vital for *T. erythraeum* survival. Inhibition of this gene expression could allow for research on the importance of trichome morphology to basic cellular processes in *T. erythraeum* including nitrogen fixation. Because prokaryotes do not have RNAi machinery, silencing of gene expression using this methodology is still not possible thus the construction of a knockout is necessary for achieving expression deficient mutants (Rusk, 2012). More research must be done to increase transformation efficiency and perform transformation on a larger scale before DNA incorporation into the genome can be carried out routinely to create recombinant *T. erythraeum* strains.

Chapter V. Conclusions

V.I. Dissertation Summary

This dissertation investigated the function of a collagen protein in the filamentous cyanobacterium, *T. erythraeum*. In chapter II, the expression and localization of this collagen protein, TrpA, revealed. TrpA is expressed through all growth phases in *T. erythraeum* however gene expression is significantly increased at Day 14 which reflects the active growth log phase in the cultures in our lab. Protein expression of TrpA is highest at Days 14 and 21. Most interestingly, using scanning laser confocal microscopy, localization of TrpA was identified in the septa between adjacent cells along the trichome. Collagenase treatment of *T. erythraeum* trichomes resulted in the liberation of cells from the filament indicating that this collagen protein plays a role in maintaining structural integrity of the trichomes by aiding in cell to cell adhesions. Being that a structural role is a characteristic of the collagen family of proteins, chapter III investigated what classification this collagen protein most closely associates with to indicate the likely structure TrpA is forming. Phylogenetic analysis indicated that TrpA most closely associates with non-fibril collagens though invertebrate fibrillar collagens were also classified in this section. Circular dichroism, protein prediction and homology modeling indicate that the glycine insertion found in this TrpA amino acid sequence introduces a perturbation that creates instability in the structure. Without electron microscopic observation of the collagen protein itself, all evidence indicates that TrpA is forming a non-fibril forming collagen. In an attempt to create a TrpA knockout strain, Chapter IV provides the first evidence that *T. erythraeum* is likely naturally transformable. Not only does the *T. erythraeum* genome contains and express a majority

of genes required for natural competence in model competent bacteria, but it has also been shown that *T. erythraeum* transformed with ampicillin resistant pUC19 survived in the presence of ampicillin.

V.II Unique presence of collagen protein in *T. erythraeum*

Currently, there are few collagen and collagen-like proteins that have been identified in prokaryotes. This could be directly linked to the fact that only recently proteins involved in bacterial cell adhesion have been identified, particularly Microbial Surface Components Recognizing Adhesive Matrix Molecules (MSCRAMM). MSCRAMMs interact with known extracellular matrix proteins in animal tissues for the attachment of colonizing microorganisms (Chagnot et al., 2012). The classification of the largest collagen sequence to date found in *T. erythraeum* is significant because it expands upon the current knowledge of the collagen family of proteins and potential impacts in the world of biopolymers (Layton et al., 2008).

Collagen-like proteins found in *Streptococcus* and *Bacillus* have been characterized and their function has been elucidated (Lukomski et al., 2000, Sylvestre et al., 2002). The Scl1 is expressed on the surface of Group A *Streptococcus* cells and has been shown to interact with human integrin α1β2. This interactions confirms that the Scl1 protein is responsible for adherence of this pathogen to the human host epithelial cells (Caswell *et al.*, 2008). Similarly, the BclA protein is expressed on the surface of the exosporium of *Bacillus anthracis.* Inactivation of *bclA* caused the loss of the filamentous appendages and identified the specific structure that the BclA protein contributes to in

these bacterial cells (Sylvestre et al., 2002). Although the specific interaction between the exosporium and human cells has yet to be elucidated, it is the first structure to interact with the host. This, indicates, again, the involvement of a prokaryotic collagen-like protein in adhesion.

Interestingly, the TrpA protein expressed in *T. erythraeum* has been localized to both the surface of trichomes and the septum between cells along the filament. The results from the collagenase treatment of trichomes indicates that this collagen protein is involved in cell adhesion of the adjacent cells along the filament to each other. Based on the research done on previous prokaryotic collagens in which the expression of the protein is found on the surface, it is highly probable that the TrpA protein found on the surface of the trichomes is involved in adhesion of trichomes to each other aiding in the maintenance of colony morphology (Figure I.3). It is also important to note that the genome of *T. erythraeum* encodes a number of integrin subunits which could be involved in the adherence of these cells and trichomes to one another through interactions mediated by TrpA.

It is also well known that *Trichodesmium* species are almost impossible to maintain in axenic cultures in the laboratory due to the presence of heterotrophic bacteria which may indicate a mutualistic relationship between the heterotrophic bacteria and *T. erythraeum* (Siddiqui *et al.*, 1992, Nausch, 1996). In these studies, electron microscopy was used to confirm the association of heterotrophic bacteria in *Trichodesmium* samples. 454 genome sequencing of the *T. erythraeum* unialgal cultures in our lab has identified the presence of at least 20 associated marine heterotrophic bacteria. The presence of heterotrophic bacteria adhering to the surface of *T. erythraeum* points to the importance

of the expression of a protein (TrpA) that would allow and enhance adhesion and consequent contact of other cells with *T. erythraeum*.

The distinction between these prokaryotic collagen-like proteins mentioned above and the TrpA protein found in *T. erythraeum* is that the collagen protein encoded in the *T. erythraeum* genome is more similar in sequence to classical collagen proteins found in eukaryotes as compared to the collagen-like proteins. They are referred to as collagen-like proteins because their collagenous-like domains exhibit residue content and distribution that varies greatly from vertebrate collagens (Ghosh *et al.*, 2012). The Scl1 collagen-like protein is 348 amino acids in length with 50 continuous Gly-x-y repeats (Lukomski et al., 2000). The BclA protein is 362 amino acids in length with 70 Gly-x-y repeats (Sylvestre et al., 2002). Both collagen-like protein sequences are significantly shorter than the TrpA protein. TrpA is 1873 amino acids in length with a collagenous domain of 558 continuous Gly-x-y repeats (Appendix A).

The TrpA sequence also reveals an extended N and C terminal region that is a common characteristic of collagen proteins. The collagenous domain is usually the majority of the amino acid sequence of fibril-forming collagens of about 300 nm in length corresponding to about 1000 amino acids in length (Gelse et al., 2003). In comparison, the collagenous domains of most other collagen types are significantly shorter, with less than 500 amino acids within the collagenous domains (Gelse et al., 2003).

On a compositional level, collagen sequences contain a glycine at every third position of the triglycine repeat with the x and y positions being any amino acid (Gly-x-y), although primarily proline is found in the one of the two positions. Eukaryotic

collagens, typically vertebrate collagens contain a hydroxyproline in the y position. This was previously believed to be a prerequisite for the stability of the collagen triple helix because of the sterochemical restrictions of the amino acid side chains as well as playing a role in creating an ordered water shell around the triple helix (Brodsky & Ramshaw, 1997). However, though bacteria lack the enzyme that post-translationally modifies proline, prolyl hydroxylase, it has recently been shown that bacterial collagen triple helices are stable in the absence of hydroxyproline (Mohs et al., 2007). Thus the lack of prolyl hydroxylase in *T. erythraeum* will not prevent the formation of collagen triple helices.

The presence of a single interruption in the collagenous domain of TrpA at residue 729 differentiates TrpA from fibrillar collagens although it is more similar in length because a prerequisite for fibrillar collagens is an uninterrupted sequence. Circular dichroism and protein prediction and modeling done in this study (chapter III) indicate the presence of the glycine interruption in TrpA does create a perturbation in the structure that decreases overall protein stability. This instability, however does not effect the formation of the collagen triple helix. The presence of this interruption is likely the main factor that influences the relatedness of the TrpA protein with non-fibrillar collagens.

Phylogenetic analysis performed as part of this investigation (Chapter III) also revealed that TrpA is closely related to invertebrate fibrillar collagens. Invertebrate fibrillar collagens commonly contain glycine imperfections in their sequence (Exposito et al., 2002). Though based on sequence analysis alone, these sequences would deem the invertebrate collagen proteins non-fibrillar collagens, isolation and electron microscopy with rotary shadowing confirm that these invertebrate sequences with interruptions in the

near perfect sequences form fibrillar collagens (Deutzmann *et al.*, 2000, Pallela *et al.*, 2011). Until isolation and rotary shadowing observations of TrpA can be made, all evidence suggests that TrpA forms a non-fibrillar collagen triple helix.

V.III TrpA and its implications in multicellularity

The structural integrity of multicellular organisms depends on the establishment and maintenance of stable cellular connections (Abedin & King, 2010). Multicellularity can develop through division without complete cell separation or through cell aggregation (Abedin & King, 2010). In either case, the transition from single cell structure to multicellularity is support by the establishment and maintenance of stable attachments between adjacent cells (Fairclough et al., 2010, Abedin & King, 2010).

The choanoflagellate, *Salpingoeca rosetta*, generates colonies that arise from the repeated division of a single cell where the daughter cells remain stably attached together to form the resulting colony (Fairclough et al., 2010). Colonial formation through aggregation has been shown to occur in many organisms including the unicellular eukaryote *Dictyostelium discoideum* through interactions of cell adhesion proteins (Abedin & King, 2010). Cell-cell adhesion molecules are expressed establishing cell-cell contacts between the unicellular cells creating an aggregate of 20-1000 cells upon nutrient stress conditions (Siu *et al.*, 2004).

Genes associated with cell-cell adhesion and multicellularity have been unexpectedly identified in a number of unicellular and colonial organisms, both prokaryotic and eukaryotic. Studies on choanoflagellates, the closest unicellular

organisms to humans, indicate that these organisms contain various proteins involved in cell adhesion including cadherin and integrin, and have been essential in understanding the evolution of multicellularity (Ruiz-Trillo et al., 2007, Fairclough et al., 2010). Despite similarities in the biological functionality of cell-cell adhesion, the specific proteins and mechanism can vary greatly between multicellular organisms (Abedin & King, 2010b).

As previously indicated, filamentous growth and elongation in *T. erythraeum* occurs through intercalary cell division, where daughter cells are attached to one another end to end after division (Stanier & Bazine, 1977). The TrpA protein localizes between these neighboring cells in *T. erythraeum* in a similar fashion as the SepJ protein in *Anabaena* PCC 7120. Inactivation mutants of SepJ and the filament integrity protein FraC show filament fragmentation in *Anabaena* PCC 7120 . These data indicate that not only are there proteins involved in maintenance of filament structure, but also that there are likely multiple proteins involved in the formation of structures that support these cell contacts in Cyanobacteria (Flores et al., 2007). Thin structures termed 'microplasmodesmata' have been identified perpendicular to the cytoplasmic membrane in the septa of the filament in *Anabaena* PCC 7120 that could be composed of the SepJ and/or FraC protein (Flores & Herrero, 2010). While these microplasmodesmata have yet to have been identified in *Trichodesmium*, a homolog of the SepJ protein has also been identified in the *T. erythraeum* genome and the TrpA protein could be involved in the formation of similar cell-cell structures.

Colony morphology is a distinct characteristic of *Trichodesmium* and was originally a standard for classification. The different *Trichodesmium* species exhibit a multitude of colony morphologies including radially-oriented puffs, parallel rafts/tufts,

and bowtie formations that are formed from the associations of up to hundreds of trichomes (Paerl & Bebout, 1992) (Fig I.3). Similar to colony formation in other organisms, colony aggregation is *Trichodesmium* species would require the presences of cell adhesion molecules that would support and maintain contact between trichomes.

V.IV Significance of a cell adhesion protein in filament integrity and nitrogen fixation

The involvement of the TrpA protein in maintenance of cell-to-cell adhesion in cells along the trichome as well as its likely involvement in adhesion of trichomes to each other suggests that is has a second function in this organism. The structural role of this protein also supports numerous mechanisms allowing for nitrogen fixation under aerobic conditions in *T. erythraeum*. Nitrogenase is a highly conserved enzyme composed of two components, an iron protein (Fe-protein) and an iron-molybdenum protein (MoFe-protein) (Berman-Frank et al., 2003a). Both the Fe_4S_4 iron-sulphur cluster found in the Fe-protein and the Fe_8S_7 cluster found in the MoFe-protein are inactivated by oxygen rendering nitrogenase unable to fix nitrogen (Berman-Frank et al., 2003). Cyanobacteria have developed structural and other adaptations that allow for oxygen evolving photosynthetic activity and nitrogen fixation while avoiding inactivation of nitrogenase by oxygen. These adaptations include temporal segregation, spatial segregation and creation of anaerobic environments.

Unicellular and most non-heterocystous filamentous diazotrophs fix nitrogen during the dark phase of the 24 hour dark-light cycle as opposed to the photoperiod to

avoid oxygen inactivation of nitrogenase. Circadian rhythm has been implicated as crucial for controlling the daily periodicity of nitrogen fixation and oxygen evolution (Zehr, 2011). It has been shown that although the two processes cannot directly co-occur, nitrogen fixation relies on ATP and electrons produced from photosynthesis (Berman-Frank et al., 2003). The unicellular cyanobacterium *Gloeothece* exhibits significant nitrogenase activity during the dark period and is supported by catabolic products accumulated during the light period (Scherer *et al.*, 1988). Similar temporal segregation has also been observed in nitrogen fixing *Synechococcus* species and filamentous nonheterocystous *Plectonema boryanum.*

The non-heterocystous filamentous cyanobacteria *Microcoleus chthonoplastes* has exhibited the ability to fix nitrogen under aerobic conditions. Researchers have implicated the tendency of this species to form bundles and aggregates in the creation of anaerobic zones (Fay, 1992). Physical structuring of cyanobacterial bundles and aggregates are important for optimizing nitrogen fixation (Paerl, 1985). The aggregates provide sites for the establishment of oxygen depleted microzones that are required for nitrogen fixation (Paerl, 1985).

In the absence of combined nitrogen, vegetative cells develop into heterocysts in a number of filamentous cyanobacteria including *Anabaena* and *Nostoc* (Adams, 2000). Heterocysts have thicker cell envelopes that aid in creating a microoxic environment preventing inactivation of nitrogenase by oxygen (Kumar *et al.*, 2010). The heterocysts and neighboring vegetative cells are interdependent in that the heterocysts require fixed carbon from the photosynthetically active cells and the vegetative cells require the fixed nitrogen from the heterocysts (Kumar et al., 2010). Structures connecting these cells are

necessary for the intercellular exchange of these metabolites (Flores & Herrero, 2010). Recently, structures termed "microplasmadesmata" have been identified in *Anabaena* PCC7120. Exchange of a 63 Da molecule has been shows to occur between cells along the filament supporting that there are connections between the cytoplasm of neighboring cells (Mullineaux *et al.*, 2008).

Nitrogen fixation during the active photoperiod is the pinnacle of research interest in *Trichodesmium erythraeum. T. erythraeum* is unique from all other filamentous nitrogen fixing cyanobacteria because it has the ability to fix nitrogen during the light phase without the presence of heterocysts. These filamentous cyanobacteria employ a number of strategies to avoid oxygen inactivation of nitrogenase including temporal segregation, spatial segregation and creation of anoxic environments. Filament integrity is critical for each of these adaptations that allow for nitrogen fixation under aerobic conditions.

While *T. erythreaum* does not exhibit nitrogen fixation during the dark phase, temporal segregation of photosynthetic activity and nitrogen fixation has been observed to occur during the photoperiod. The core protein, D1, of the oxygen-evolving PSII reaction center is found to colocalize to all cells along the trichome including those containing nitrogenase. Photosynthetic carbon fixation increases in the morning, declines midday and then increases again at the end at the photoperiod, while nitrogen fixation was shown to peak midday when photosynthetic activity exhibits a decline (Berman-Frank et al., 2001). This cycle of photosynthetic activity correlates with *Trichodesmium* buoyancy regulation in the water column. To avoid photo damage, after periods of high irradiance, gas vesicles collapse and decrease *Trichodesmium* positioning in the water

column which in turn decreases photosynthetic activity and increases nitrogen fixation because of the decreased light penetration in the water column. Consequently, the gas vesicles again expand and positioning them higher in the water column closer to the surface which increases exposure to light and increases photosynthetic activity (Walsby, 1994, Walsby, 1992). Trichome structure is a prerequisite for colony formation and these colonies are critical to increasing the rate at which buoyancy regulation occurs (Walsby, 1992).

Although *Trichodesmium* spp. do not exhibit heterocyst differentiation, recent studies have shown that a subset of cells along a trichome exhibit increased levels of respiratory enzymes and nitrogenase similar to heterocysts (Sandh *et al.*, 2012). These subset of cells, diazocytes, are observed to be regularly spaced between chains of vegetative cells similar to that observed in heterocystous cyanobacteria. Additionally, the hetR gene that signals for heterocyst development is also encoded in *Trichodesmium*, which may be signaling for the induction of diazocytes under nitrogen stress (Sandh et al., 2012). Although there has yet to have been any research to indicate that there are junctions connecting the cytoplasm of adjacent cells in *T. erythraeum* it is likely that intercellular exchange of these metabolites would occur similarly to that of the filamentous heterocystous diazotrophs. Again, trichome integrity is required for this mechanism of nitrogen fixation to be sustainable because it would be necessary for the cell-to-cell adhesion to be maintained for the vegetative cells to receive fixed nitrogen and the diazocytes to receive fixed carbon.

Anaerobic microzones have been long implicated as an adaptation in *Trichodesmium* that allows for nitrogen fixation under aerobic conditions. Self-shading

occurs with colony and bloom formation decreasing absorbance (Prufert-Bebout et al., 1993, Subramaniam et al., 1999). Self-shading that occurs during aggregation would create areas of colonies or blooms that decreases photosynthetic activity thus creating zones that have decreased oxygen evolution (Carpenter *et al.*, 2004). These low-oxygen zones created by colony formation are crucial in regulation of nitrogen fixation (Paerl *et al.*, 1989). For the formation of colonies, maintenance of cells-to-cell contact along the filament and maintenance of contact between trichomes is required to withstand separation and fragmentation during natural mixing of the eutropic zone.

It is clear that a number of strategies are employed to allow nitrogen fixation to occur in *T. erythraeum* under aerobic conditions. As the nitrogenase enzyme itself is not structurally different to prevent inactivation from oxygen, morphological adaptations of *T. erythraeum* are critical. Though the mechanisms to prevent the inactivation of nitrogenase by oxygen vary, the common factor in each is the importance of trichomes to resist mechanical stress both along the trichome itself and between other trichomes.

V.V TrpA as a biopolymer

Based on its structural and mechanical properties presented in this work, TrpA can provide biomaterial researchers with a source of collagen that can be used in various applications of collagen engineering. Many characteristics of collagen, including biocompatibility and weak antigenicity, make it a desirable biomaterial for medical applications (Lee *et al.*, 2001). It is more desirable than synthetic polymers because it is absorbable in the body and has a high affinity with water (Miyata *et al.*, 1992).

Collagen has been used as a mechanism for drug and gene delivery in the form of films, matrix systems, collagen shields, pellets and nanoparticles. It is also the basis for tissue engineering helping with skin replacement, bone substitution and tissue development (Lee et al., 2001). Skin is the usual source for large quantities of collagen. However, this is not the most desirable source of collagen because the tissue source must be completely sterile and free of any microorganism at the initiation of collagen extraction to assure sterility of the collagen (Miyata et al., 1992).

Potential disease transmission and immunorejection of commercially available collagen has driven the pursuit of alternate collagen sources (Cunningham *et al.*, 2010). Recently, marine sponge collagen has been used to create a scaffold for bone tissue engineering (Pallela et al., 2011). It has been demonstrated that marine sponge collagen is an acceptable tissue to use for tissue engineering; more so than synthetic constructs (Green 2008). This marine source of collagen was of interest because the soft collagenous skeleton was composed of both fibrillar and non-fibrillar collagen (Pallela et al., 2011).

Based on this study of TrpA, it is clear that it is capable of forming a collagen molecule. This research is significant because it reveals a prokaryotic source of collagen that can be investigated for bioengineering. This source of collagen would provide a less expensive and non-infectious source that could be used for the engineering of collagen biomaterials including nanoparticles, minipellets, and sponges for drug delivery (Senni *et al.*, 2011, Lee et al., 2001). Additionally, since TrpA is not from an animal, there is less likelihood of an immune reaction from the host.

V.VI Conclusion and Future studies

The research presented in this dissertation clearly indicates that the collagen protein, TrpA, expressed in *T. erythraeum* functions as a structural protein similar to that of all other members belonging to the collagen family of proteins. It would be of particular interest to preserve colony morphology followed by immunolabeling of trichomes to conclusively determine the role collagen plays in the adhesion of trichomes to one another. The achievement of a TrpA knockout would be invaluable to determining if the absence of the collagen protein would have any additional structural effects on trichomes. For instance, would a TrpA mutant strain result in a unicellular morphology with liberation of the daughter cells after cell division as opposed to filament formation or would this be a lethal mutant?

Isolation of the collagen molecule expressed in *T. erythraeum* would be vital to proposing a mechanism of adhesion between cells along the filament as well as trichomes within a colony. It would be of importance to use electron microscopy with rotary shadowing to identify the suprastructure that is formed by TrpA. This would answer questions such as does TrpA form fibrils, does cleavage of the N- and/or C-terminal occur, or does the glycine insertion create turns in the collagen molecule? Though circular dichroism is a good indicator for fold and stability of collagen peptides, isolation and subsequent crystallization would provide a higher level of structural detail (Fallas *et al.*, 2009, Berisio *et al.*, 2002). Resolving the structure would be critical in moving forward towards utilizing TrpA as a collagen biopolymer.

References

Abedin, M. & King, N. 2010. Diverse evolutionary paths to cell adhesion. *Trends in cell biology* **20**:734-42.

Adams, D. G. 2000. Heterocyst formation in cyanobacteria. *Current Opinion in Microbiology* **3**:618-24.

Anandan, S. & Uram, J. 2004. Genetic selection scheme for isolation of signal transduction pathway mutants. *Applied and Environmental Microbiology* **70**:967-72.

Andersen, R. A. 2005. *Algal Culturing Techniques.* Academic Press, San Diego

Ando, T., Xu, Q., Torres, M., Kusugami, K., Israel, D. A. & Blaser, M. J. 2000. Restriction-modification system differences in *Helicobacter pylori* are a barrier to interstrain plasmid transfer. *Molecular Microbiology* **37**:1052-65.

Andresen, E., Lohscheider, J., Šetlikova, E., Adamska, I., Šimek, M. & Küpper, H. 2010. Acclimation of *Trichodesmium erythraeum* ISM101 to high and low irradiance analysed on the physiological, biophysical and biochemical level. *New Phytologist* **185**:173-88.

Arnold, K., Bordoli, L., Kopp, J. r. & Schwede, T. 2006. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* **22**:195-201.

Bächinger, H. P., Morris, N., Lunstrum, G. P., Keene, D. R., Rosenbaum, L. M., Compton, L. A. & Burgeson, R. 1990. Relationship of the Biophysical and Biochemical Characteristics of Type VII Collagen to the Function of Anchoring Fibrils. *The Journal of Biological Chemistry* **265**:10095-101.

Bella, J., Liu, J., Kramer, R., Brodsky, B. & Berman, H. M. 2006. Conformational Effects of Gly-X-Gly Interruptions in the Collagen Triple Helix. *Journal of molecular biology* **362**:298-311.

Bergman, B., Sandh, G., Lin, S., Larsson, J. & Carpenter, E. J. 2012. *Trichodesmium* - a widespread marine cyanobacterium with unusual nitrogen fixation properties. *FEMS Microbiology Reviews*.

Berisio, R., Vitagliano, L., Mazzarella, L. & Zagari, A. 2002. Crystal structure of the collagen triple helix model [(Pro-Pro-Gly) 10] 3. *Protein science* **11**:262-70.

Berman-Frank, I., Lundgren, P., Chen, Y.-B., Küpper, H., Kolber, Z., Bergman, B. & Falkowski, P. 2001. Segregation of nitrogen fixation and oxygenic photosynthesis in the marine cyanobacterium *Trichodesmium*. *Science* **294**:1534-37.

Berman-Frank, I., Lundgren, P. & Falkowski, P. 2003. Nitrogen fixation and photosynthetic oxygen evolution in cyanobacteria. *Research in Microbiology* **154**:157-64.

Boot-Handford, R. P. & Tuckwell, D. S. 2003. Fibrillar collagen: the key to vertebrate evolution? A tale of molecular incest. *Bioessays* **25**:142-51.

Bordoli, L., Kiefer, F., Arnold, K., Benkert, P., Battey, J. & Schwede, T. 2008. Protein structure homology modeling using SWISS-MODEL workspace. *Nature Protocols* **4**:1-13.

Borstad, G. A. & Borstad, L. E. 1984. Needlelike projectiles observed issuing from lysing *Trichodesmium* (Cyanophyta) cells. *Limnology and Oceanography* **29**:1097-99.

Boudko, S. P., Engel, J. r. & Bächinger, H. P. 2012. The crucial role of trimerization domains in collagen folding. *The International Journal of Biochemistry & Cell Biology* **44**:21-32.

Boydston, J. A., Chen, P., Steichen, C. T. & Turnbough, C. L., Jr. 2005. Orientation within the exosporium and structural stability of the collagen-like glycoprotein BclA of *Bacillus anthracis*. *Journal of bacteriology* **187**:5310-7.

Brahmbhatt, T. N., Janes, B. K., Stibitz, E. S., Darnell, S. C., Sanz, P., Rasmussen, S. B. & O'Brien, A. D. 2007. *Bacillus anthracis* exosporium protein BclA affects spore germination, interaction with extracellular matrix proteins, and hydrophobicity. *Infection and Immunity* **75**:5233-9.

Brodsky, B. & Ramshaw, J. A. 1997. The collagen triple-helix structure. *Matrix Biology* **15**:545-54.

Burgeson, R. 1993. Type VII Collagen, Anchoring Fibrils, and Epidermolysis Bullosa. *The Journal for Investigative Dermatology* **101**:252-55.

Capone, D., Zehr, J., Paerl, H., Bergman, B. & Carpenter, E. 1997. Trichodesmium, a Globally Significant Marine Cyanobacterium. *Science* **276**:1221-29.

Carpenter, E. J., Subramaniam, A. & Capone, D. G. 2004. Biomass and primary productivity of the cyanobacterium *Trichodesmium* spp. in the tropical N Atlantic ocean. *Deep Sea Research Part I: Oceanographic Research Papers* **51**:173-203.

Caswell, C. C., Han, R., Hovis, K. M., Ciborowski, P., Keene, D. R., Marconi, R. T. & Lukomski, S. 2008. The Scl1 protein of M6-type group A *Streptococcus* binds the human complement regulatory protein, factor H, and inhibits the alternative pathway of complement. *Molecular Microbiology* **67**:584-96.

Chagnot, C., Listrat, A., Astruc, T. & Desvaux, M. 2012. Bacterial adhesion to animal tissues: protein determinants for recognition of extracellular matrix components. *Cellular Microbiology* **14**:1687-96.

Chen, I. S., Christie, P. J. & Dubnau, D. 2005. The ins and outs of DNA transfer in bacteria. *Science* **310**:1456-60.

Chen, I. S. & Dubnau, D. 2004. DNA uptake during bacterial transformation. *Nature Reviews Microbiology* **2**:241-49.

Chen, Y.-B., Zehr, J. P. & Mellon, M. 1996. Growth and nitrogen fixation of the diazotrophic filamentous nonheterocystous cyanobacterium *Trichodesmium* sp. IMS 101 in defined media: evidence for a circadian rhythm. *Journal of Phycology* **32**:916-23.

Chiang, G., Schaefer, M. & Grossman, A. 1992. Transformation of the filamentous cyanobacterium Fremyella diplosiphon by conjugaison or electroporation. *Plant Physiology and Biochemistry* **30**:315-25.

Claverys, J.-P. & Martin, B. 2003. Bacterial 'competence' genes: signatures of active transformation, or only remnants? *Trends in Microbiology* **11**:161-65.

Coelho, S. M., Simon, N., Ahmed, S., Cock, J. M. & Partensky, F. 2013. Ecological and evolutionary genomics of marine photosynthetic organisms. *Molecular Ecology* **22**:867-907.

Cunningham, E., Dunne, N., Walker, G., Maggs, C., Wilcox, R. & Buchanan, F. 2010. Hydroxyapatite bone substitutes developed via replication of natural marine sponges. *Journal of Materials Science: Materials in Medicine* **21**:2255-61.

Deutzmann, R., Fowler, S., Zhang, X., Boone, K., Dexter, S., Boot-Handford, R., Rachel, R. & Sarras, M. 2000. Molecular, biochemical and functional analysis of a novel and developmentally important fibrillar collagen (Hcol-I) in hydra. *Development* **127**:4669-80.

El-Shehawy, R., Lugomela, C., Ernst, A. & Bergman, B. 2003. Diurnal expression of hetR and diazocyte development in the filamentous non-heterocystous cyanobacterium *Trichodesmium erythraeum*. *Microbiology* **149**:1139-46.

Elhai, J. 1994. Genetic techniques appropriate for the biotechnological exploitation of cyanobacteria. *Journal of Applied Phycology* **6**:177-86.

Espie, G. S. & Kimber, M. S. 2011. Carboxysomes: cyanobacterial RubisCO comes in small packages. *Photosynthesis research* **109**:7-20.

Exposito, J.-Y. & Garrone, R. 1990. Characterization of a fibrillar collagen gene in sponges reveals the early evolutionary appearance of two collagen gene families. *Proceedings of the National Academy of Sciences* **87**:6669-73.

Exposito, J.-Y., Larroux, C., Cluzel, C., Valcourt, U., Lethias, C. & Degnan, B. M. 2008. Demosponge and sea anemone fibrillar collagen diversity reveals the early emergence of A/C clades and the maintenance of the modular structure of type V/XI collagens from sponge to human. *Journal of Biological Chemistry* **283**:28226-35.

Exposito, J. Ä., Cluzel, C., Garrone, R. & Lethias, C. 2002. Evolution of collagens. *The Anatomical Record* **268**:302-16.

Fairclough, S. R., Dayel, M. J. & King, N. 2010. Multicellular development in a choanoflagellate. *Curr Biol* **20**:R875-6.

Fallas, J. A., Gauba, V. & Hartgerink, J. D. 2009. Solution structure of an ABC collagen heterotrimer reveals a single-register helix stabilized by electrostatic interactions. *Journal of Biological Chemistry* **284**:26851-59.

Fay, P. 1992. Oxygen relations of nitrogen fixation in cyanobacteria. *Microbiological Reviews* **56**:340.

Finkel, S. E. & Kolter, R. 2001. DNA as a nutrient: novel role for bacterial competence gene homologs. *Journal of bacteriology* **183**:6288-93.

Flores, E. & Herrero, A. 2010. Compartmentalized function through cell differentiation in filamentous cyanobacteria. *Nature Reviews Microbiology* **8**:39-50.

Flores, E., Muro-Pastor, A., Meeks, J., Herrero, A. & Flores, E. 2008. Gene transfer to cyanobacteria in the laboratory and in nature. *The Cyanobacteria: Molecular Biology, Genomics and Evolution*:45-57.

Flores, E., Pernil, R., Muro-Pastor, A. M., Mariscal, V., Maldener, I., Lechno-Yossef, S., Fan, Q., Wolk, C. P. & Herrero, A. 2007. Septum-localized protein required for filament integrity and diazotrophy in the heterocyst-forming cyanobacterium *Anabaena* sp. strain PCC 7120. *Journal of bacteriology* **189**:3884-90.

Flores, E. & Wolk, C. P. 1985. Identification of facultatively heterotrophic, N2-fixing cyanobacteria able to receive plasmid vectors from Escherichia Coli by conjugation. *Journal of bacteriology* **162**:1339-41.

Gantt, E. 1975. Phycobilisomes: Light-harvesting pigment complexes. *Bioscience*:781-88.

Geer, L. Y., Marchler-Bauer, A., Geer, R. C., Han, L., He, J., He, S., Liu, C., Shi, W. & Bryant, S. H. 2010. The NCBI BioSystems database. *Nucleic acids research* **38**:D492-D96.

Gelse, K., Pöschl, E. & Aigner, T. 2003. Collagens--structure, function, and biosynthesis. *Advanced Drug Delivery Reviews* **55**:1531-46.

Ghosh, N., McKillop, T. J., Jowitt, T. A., Howard, M., Davies, H., Holmes, D. F., Roberts, I. S. & Bella, J. 2012. Collagen-Like Proteins in Pathogenic *E. coli* Strains. *PloS one* **7**:e37872.

Golden, S. S., Brusslan, J. & Haselkorn, R. 1987. Genetic engineering of the cyanobacterial chromosome. *Methods Enzymol.* **153**:215-31.

Golden, S. S. & Sherman, L. A. 1984. Optimal conditions for genetic transformation of the cyanobacterium *Anacystis nidulans* R2. *Journal of Bacteriology* **158**:36-42.

Gouy, M., Guindon, S. & Gascuel, O. 2010. SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Molecular Biology and Evolution* **27**:221-24.

Guex, N. & Peitsch, M. C. 1997. SWISS-MODEL and the Swiss-Pdb Viewer: An environment for comparative protein modeling. *Electrophoresis* **18**:2714-23.

Guindon, S., Lethiec, F., Duroux, P. & Gascuel, O. 2005. PHYML Online--a web server for fast maximum likelihood-based phylogenetic inference. *Nucleic Acids Research* **33**:W557-W59.

Gupta, R. S. 2003. Evolutionary relationships among photosynthetic bacteria. *Photosynthesis research* **76**:173-83.

Hewson, I., Poretsky, R. S., Dyhrman, S. T., Zielinski, B., White, A. E., Tripp, H. J., Montoya, J. P. & Zehr, J. P. 2009. Microbial community gene expression within colonies of the diazotroph, *Trichodesmium*, from the Southwest Pacific Ocean. *The ISME journal* **3**:1286-300.

Hulmes, D. J. 2002. Building collagen molecules, fibrils, and suprafibrillar structures. *J Struct Biol* **137**:2-10.

Hwang, E. S. & Brodsky, B. 2012. Folding delay and structural perturbations caused by type IV collagen natural interruptions and nearby Gly missense mutations. *Journal of Biological Chemistry* **287**:4368-75.

Hynes, A. M., Webb, E. A., Doney, S. C. & Waterbury, J. B. 2012. Comparison of cultured *Trichodesmium* (Cyanophyceae) with species characterized from the field. *Journal of Phycology* **48**:196-210.

Ives, C., Nathan, P. & Brooks, J. 1992. Regulation of the BamHI restriction-modification system by a small intergenic open reading frame, bamHIC, in both *Escherichia coli* and *Bacillus subtilis*. *Journal of Bacteriology* **174**:7194-201.

Janson, S., Siddiqui, P. J., Walsby, A. E., Romans, K. M., Carpenter, E. J. & Bergman, B. 1995. Cytomorphological characterization of the planktonic diazotrophic cyanobacteria *Trichodesmium* spp. From the Indian Ocean and Caribbean and Sargasso Seas. *Journal of Phycology* **31**:463-77.

Johnsborg, O., Eldholm, V. & Håvarstein, L. S. 2007. Natural genetic transformation: prevalence, mechanisms and function. *Research in Microbiology* **158**:767-78.

Kadler, K. E., Baldock, C., Bella, J. & Boot-Handford, R. P. 2007. Collagens at a glance. *Journal of Cell Science* **120**:1955-58.

Karlstrom, A., Jacobsson, K., Flock, M., Flock, J. I. & Guss, B. 2004. Identification of a novel collagen-like protein, SclC, in *Streptococcus equi* using signal sequence phage display. *Vet Microbiol* **104**:179-88.

Kasting, J. F. 1993. Earth's early atmosphere. *Science* **259**:920-26.

Kielty, C. & Grant, M. 2003. The collagen family: structure, assembly, and organization in the extracellular matrix. *In*: Royce, P. M. & Steinmann, B. [Eds.] *Connective Tissue and Its Heritable Disorders: Molecular, Genetic, and Medical Aspects*. Second ed. John Wiley & Sons, Inc., Hoboken pp. 159-221.

King, N., Westbrook, M. J., Young, S. L., Kuo, A., Abedin, M., Chapman, J., Fairclough, S., Hellsten, U., Isogai, Y., Letunic, I., Marr, M., Pincus, D., Putnam, N., Rokas, A., Wright, K. J., Zuzow, R., Dirks, W., Good, M., Goodstein, D., Lemons, D., Li, W., Lyons, J. B., Morris, A., Nichols, S., Richter, D. J., Salamov, A., Sequencing, J. G., Bork, P., Lim, W. A., Manning, G., Miller, W. T., McGinnis, W., Shapiro, H., Tjian, R., Grigoriev, I. V. & Rokhsar, D. 2008. The genome of the choanoflagellate *Monosiga brevicollis* and the origin of metazoans. *Nature* **451**:783-8.

Klint, J., Rasmussen, U. & Bergman, B. 2007. FtsZ may have dual roles in the filamentous cyanobacterium *Nostoc/Anabaena* sp. strain PCC 7120. *Journal of Plant Physiology* **164**:11-8.

Klock, G., Kowalski, M. B., Hering, B. J., Eiden, M. E., Weidemann, A., Langer, S., Zimmermann, U., Federlin, K. & Bretzel, R. G. 1996. Fractions from commercial collagenase preparations: use in enzymic isolation of the islets of Langerhans from porcine pancreas. *Cell Transplant* **5**:543-51.

Koksharova, O. & Wolk, C. 2002. Genetic tools for cyanobacteria. *Applied microbiology and Biotechnology* **58**:123-37.

Kondo, T., Strayer, C. A., Kulkarni, R. D., Taylor, W., Ishiura, M., Golden, S. S. & Johnson, C. H. 1993. Circadian rhythms in prokaryotes: luciferase as a reporter of circadian gene expression in cyanobacteria. *Proceedings of the National Academy of Sciences* **90**:5672-76.

Krane, D. E. & Raymer, M. Fundamental Concepts of Bioinformatics. 2003. Benjamin Cummings.

Kumar, K., Mella-Herrera, R. A. & Golden, J. W. 2010. Cyanobacterial heterocysts. *Cold Spring Harbor Perspectives in Biology* **2**.

Layton, B. E., D'Souza, A. J., Dampier, W., Zeiger, A., Sabur, A. & Jean-Charles, J. 2008. Collagen's triglycine repeat number and phylogeny suggest an interdomain transfer event from a Devonian or Silurian organism into *Trichodesmium erythraeum*. *Journal of molecular evolution* **66**:539-54.

Lee, C. H., Singla, A. & Lee, Y. 2001. Biomedical applications of collagen. *International Journal of Pharmaceutics* **221**:1-22.

Letunic, I. & Bork, P. 2011. Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. *Nucleic Acids Research* **39**:W475-W78.

Lin, S., Henze, S., Lundgren, P., Bergman, B. & Carpenter, E. J. 1998. Whole-Cell Immunolocalization Of Nitrogenase In Marine Diazotrophic Cyanobacteria, *Trichodesmium* spp. *Applied and Environmental Microbiology* **64**:3052-8.

Lukomski, S., Nakashima, K., Abdi, I., Cipriano, V. J., Ireland, R. M., Reid, S. D., Adams, G. G. & Musser, J. M. 2000. Identification and Characterization of thescl Gene Encoding a Group A *Streptococcus* Extracellular Protein Virulence Factor with Similarity to Human Collagen. *Infection and immunity* **68**:6542-53.

MacColl, R. 1998. Cyanobacterial phycobilisomes. *J Struct Biol* **124**:311-34.

Mann, K., Gaill, F. & Timpl, R. 1992. Amino-acid sequence and cell-adhesion activity of a fibril-forming collagen from the tube worm *Riftia pachyptila* living at deep sea hydrothermal vents. *European Journal of Biochemistry* **210**:839-47.

Marga, F., Grandbois, M., Cosgrove, D. J. & Baskin, T. I. 2005. Cell wall extension results in the coordinate separation of parallel microfibrils: evidence from scanning electron microscopy and atomic force microscopy. *The Plant Journal* **43**:181-90.

Mariscal, V., Herrero, A., Nenninger, A., Mullineaux, C. W. & Flores, E. 2011. Functional dissection of the three-domain SepJ protein joining the cells in cyanobacterial trichomes. *Molecular Microbiology* **79**:1077-88.

Miyata, T., Taira, T. & Noishiki, Y. 1992. Collagen engineering for biomaterial use. *Clinical Materials* **9**:139-48.

Mohs, A., Silva, T., Yoshida, T., Amin, R., Lukomski, S., Inouye, M. & Brodsky, B. 2007. Mechanism of stabilization of a bacterial collagen triple helix in the absence of hydroxyproline. *Journal of Biological Chemistry* **282**:29757-65.

Monson, J., Natzle, J., Friedman, J. & McCarthy, B. 1982. Expression and novel structure of a collagen gene in *Drosophila*. *Proceedings of the National Academy of Sciences* **79**:1761-65.

Mullineaux, C. W., Mariscal, V., Nenninger, A., Khanum, H., Herrero, A., Flores, E. & Adams, D. G. 2008. Mechanism of intercellular molecular exchange in heterocyst-forming cyanobacteria. *The EMBO journal* **27**:1299-308.

Nausch, M. 1996. Microbial activities on *Trichodesmium* colonies. *Mar. Ecol. Prog. Ser. Oldendorf* **141**:173-81.

Nimni, M. E., Cheung, D., Strates, B., Kodama, M. & Sheikh, K. 1987. Chemically modified collagen: a natural biomaterial for tissue replacement. *J Biomed Mater Res* **21**:741-71.

Ohki, K. & Fujita, Y. 1982. Laboratory culture of the pelagic blue-green alga *Trichodesmium thiebautii*: Conditions for unialgal culture. *Mar. Ecol. Prog. Ser* **7**:185-90.

Orcutt, K., Rasmussen, U., Webb, E. A., Waterbury, J. B., Gundersen, K. & Bergman, B. 2002. Characterization of Trichodesmium spp. by genetic techniques. *Applied and Environmental Microbiology* **68**:2236-45.

Paerl, H. & Bebout, B. 1992. Oxygen dynamics in *Trichodesmium* spp. aggregates. *Marine Pelagic Cyanobacteria: Trichodesmium and Other Diazotrophs.* Springer, pp. 43-59.

Paerl, H. W. 1985. Microzone formation: its role in the enhancement of aquatic N2 fixation. *Limnol. Oceanogr* **30**:1246-52.

Paerl, H. W., Priscu, J. C. & Brawner, D. L. 1989. Immunochemical localization of nitrogenase in marine *Trichodesmium* aggregates: relationship to N2 fixation potential. *Applied and Environmental Microbiology* **55**:2965-75.

Pallela, R., Bojja, S. & Janapala, V. R. 2011. Biochemical and biophysical characterization of collagens of marine sponge, *Ircinia fusca* (Porifera: Demospongiae: Irciniidae). *International Journal of Biological Macromolecules* **49**:85-92.

Pêcher, J., Pires, V., Djaafri, I., Da Nascimento, S., Fauvel-Lafève, F., Legrand, C. & Sonnet, P. 2009. Circular dichroism studies of type III collagen mimetic peptides with anti-or pro-aggregant activities on human platelets. *European Journal of Medicinal Chemistry* **44**:2643-50.

Price, S. & Anandan, S. 2009. Protein Sequence Analysis. Drexel University, Philadelphia.

Prufert-Bebout, L., Paerl, H. W. & Lassen, C. 1993. Growth, nitrogen fixation, and spectral attenuation in cultivated *Trichodesmium* species. *Applied and Environmental Microbiology* **59**:1367-75.

Ramachandran, G. & Kartha, G. 1954. Structure of collagen. *Nature.*174: 269-270

Rasmussen, M., Eden, A. & Bjorck, L. 2000. SclA, a novel collagen-like surface protein of Streptococcus pyogenes. *Infection and Immunity* **68**:6370-7.

Rasmussen, M., Jacobsson, M. & Bjorck, L. 2003. Genome-based identification and analysis of collagen-related structural motifs in bacterial and viral proteins. *The Journal of Biological Chemistry* **278**:32313-6.

Raven, J. & Allen, J. 2003. Genomics and chloroplast evolution: what did cyanobacteria do for plants? *Genome Biology* **4**:209.

Ricard-Blum, S. 2011. The collagen family. *Cold Spring Harbor Perspectives In Biology* **3**:a004978.

Rippka, R., Deruelles, J., Waterbury, J. B., Herdman, M. & Stanier, R. Y. 1979. Generic assignments, strain histories and properties of pure cultures of cyanobacteria. *Journal of General Microbiology* **111**:1-61.

Rodbell, M. 1964. Localization of Lipoprotein Lipase in Fat Cells of Rat Adipose Tissue. *The Journal of Biological Chemistry* **239**:753-5.

Roy, A., Kucukural, A. & Zhang, Y. 2010. I-TASSER: a unified platform for automated protein structure and function prediction. *Nature Protocols* **5**:725-38.

Ruiz-Trillo, I., Burger, G., Holland, P. W., King, N., Lang, B. F., Roger, A. J. & Gray, M. W. 2007. The origins of multicellularity: a multi-taxon genome initiative. *Trends in Genetics* **23**:113-18.

Rusk, N. 2012. Microbiology: Prokaryotic RNAi. *Nat Methods* **9**:220-21.

Sanchez-Baracaldo, P., Hayes, P. & Blank, C. 2005. Morphological and habitat evolution in the Cyanobacteria using a compartmentalization approach. *Geobiology* **3**:145-65.

Sandh, G., Xu, L. & Bergman, B. 2012. Diazocyte development in the marine diazotrophic cyanobacterium *Trichodesmium*. *Microbiology* **158**:345-52.

Scherer, S., Almon, H. & Böger, P. 1988. Interaction of photosynthesis, respiration and nitrogen fixation in cyanobacteria. *Photosynthesis Research* **15**:95-114.

Schirrmeister, B., Antonelli, A. & Bagheri, H. 2011. The origin of multicellularity in cyanobacteria. *BMC Evolutionary Biology* **11**:45.

Schwede, T., Kopp, J. r., Guex, N. & Peitsch, M. C. 2003. SWISS-MODEL: an automated protein homology-modeling server. *Nucleic Acids Research* **31**:3381-85.

Sellner, K. 1992. Trophodynamics of marine cyanobacteria blooms. *Marine pelagic cyanobacteria: Trichodesmium and Other Diazotrophs.* Springer, pp. 75-94.

Senni, K., Pereira, J., Gueniche, F., Delbarre-Ladrat, C., Sinquin, C., Ratiskol, J., Godeau, G., Fischer, A.-M., Helley, D. & Colliec-Jouault, S. 2011. Marine polysaccharides: a source of bioactive molecules for cell therapy and tissue engineering. *Marine Drugs* **9**:1664-81.

Shih, P. M., Wu, D., Latifi, A., Axen, S. D., Fewer, D. P., Talla, E., Calteau, A., Cai, F., de Marsac, N. T. & Rippka, R. 2013. Improving the coverage of the cyanobacterial phylum using diversity-driven genome sequencing. *Proceedings of the National Academy of Sciences* **110**:1053-58.

Shoulders, M. D. & Raines, R. T. 2009. Collagen structure and stability. *Annu Rev Biochem* **78**:929.

Sicot, F.-X., Exposito, J.-Y., Masselot, M., Garrone, R., Deutsch, J. & Gaill, F. 1997. Cloning of an Annelid Fibrillar-Collagen Gene and Phylogenetic Analysis of Vertebrate and Invertebrate Collagens. *European Journal of Biochemistry* **246**:50-58.

Siddiqui, P., Bergman, B. & Carpenter, E. 1992. Filamentous cyanobacterial associates of the marine planktonic cyanobacterium *Trichodesmium*. *Phycologia* **31**:326-37.

Siu, C.-H., Harris, T. J., Wang, J. & Wong, E. 2004. Regulation of cell-cell adhesion during *Dictyostelium* development. *Seminars in Cell & Developmental Biology.* Elsevier, pp. 633-41.

Stal, L. J. 2012. *Cyanobacterial mats and stromatolites.* Springer, Netherlands 65-125

Stanier, R. & Bazine, G. 1977. Phototrophic prokaryotes: the cyanobacteria. *Annual Reviews in Microbiology* **31**:225-74.

Su, C. M., Hsueh, H. T., Chen, H. H. & Chu, H. 2012. Effects of dissolved inorganic carbon and nutrient levels on carbon fixation and properties of *Thermosynechococcus* sp. in a continuous system. *Chemosphere*.

Subramaniam, A., Brown, C. W., Hood, R. R., Carpenter, E. J. & Capone, D. G. 2002. Detecting *Trichodesmium* blooms in SeaWiFS imagery. *Deep Sea Research Part II: Topical Studies in Oceanography* **49**:107-21.

Subramaniam, A., Carpenter, E. J. & Falkowski, P. G. 1999. Bio-optical properties of the marine diazotrophic cyanobacteria *Trichodesmium* spp. II. A reflectance model for remote sensing. *Limnology and Oceanography* **44**:618-27.

Sylvestre, P., Couture-Tosi, E. & Mock, M. 2003. Polymorphism in the collagen-like region of the Bacillus anthracis BclA protein leads to variation in exosporium filament length. *Journal of Bacteriology* **185**:1555-63.

Sylvestre, P., Couture-Tosi, E. & Mock, M. 2002. A collagen-like surface glycoprotein is a structural component of the Bacillus anthracis exosporium. *Molecular Microbiology* **45**:169-78.

Taniuchi, Y., Murakami, A. & Ohki, K. 2008. Whole-cell immunocytochemical detection of nitrogenase in cyanobacteria: improved protocol for highly fluorescent cells. *Aquatic Microbial Ecology* **51**:237-47.

Thiel, T. & Poo, H. 1989. Transformation of a filamentous cyanobacterium by electroporation. *Journal of Bacteriology* **171**:5743-46.

Toal, S., Amidi, O. & Schweitzer-Stenner, R. 2011. Conformational Changes of Trialanine Induced by Direct Interactions between Alanine Residues and Alcohols in Binary Mixtures of Water with Glycerol and Ethanol. *Journal of the American Chemical Society* **133**:12728-39.

Todd, S. J., Moir, A. J., Johnson, M. J. & Moir, A. 2003. Genes of Bacillus cereus and Bacillus anthracis encoding proteins of the exosporium. *Journal Of Bacteriology* **185**:3373-8.

van Baalen, C. & Brown Jr, R. M. 1969. The ultrastructure of the marine blue green alga, *Trichodesmium erythraeum*, with special reference to the cell wall, gas vacuoles, and cylindrical bodies. *Archiv für Mikrobiologie* **69**:79-91.

Vishnyakov, I. & Borchsenius, S. 2007. FtsZ and bacterial cell division. *Cell and Tissue Biology* **1**:206-14.

Vuorio, E. & de Crombrugghe, B. 1990. The family of collagen genes. *Annu Rev Biochem* **59**:837-72.

Walsby, A. 1978. The gas vesicles of aquatic prokaryotes. *Symp. Soc. Gen. Microbiol.* pp. 327-58.

Walsby, A. 1992. The gas vesicles and buoyancy of *Trichodesmium*. *Marine Pelagic Cyanobacteria: Trichodesmium and Other Diazotrophs*. Springer, pp. 141-61.

Walsby, A. 1994. Gas vesicles. *Microbiological Reviews* **58**:94.

Waterbury, J. B. 2006. The cyanobacteria--isolation, purification and identification. *Prokaryotes* **4**:1053-73.

Webb, E. A., Moffett, J. W. & Waterbury, J. B. 2001. Iron stress in open-ocean cyanobacteria (*Synechococcus, Trichodesmium*, and *Crocosphaera* spp.): identification of the IdiA protein. *Applied and Environmental Microbiology* **67**:5444-52.

Whatmore, A. M. 2001. *Streptococcus pyogenes* sclB encodes a putative hypervariable surface protein with a collagen-like repetitive structure. *Microbiology* **147**:419-29.

Wolk, C. P., Vonshak, A., Kehoe, P. & Elhai, J. 1984. Construction of shuttle vectors capable of conjugative transfer from *Escherichia coli* to nitrogen-fixing filamentous cyanobacteria. *Proceedings of the National Academy of Sciences* **81**:1561-65.

Yeates, T. O., Kerfeld, C. A., Heinhorst, S., Cannon, G. C. & Shively, J. M. 2008. Protein-based organelles in bacteria: carboxysomes and related microcompartments. *Nature Reviews Microbiology* **6**:681-91.

Yu, Z., Mirochnitchenko, O., Xu, C., Yoshizumi, A., Brodsky, B. & Inouye, M. 2010. Noncollagenous region of the streptococcal collagen-like protein is a trimerization domain that supports refolding of adjacent homologous and heterologous collagenous domains. *Protein Sci* **19**:775-85.

Zehr, J. P. 2011. Nitrogen fixation by marine cyanobacteria. *Trends in Microbiology* **19**:162-73.

Zehr, J. P., Carpenter, E. J. & Villareal, T. A. 2000. New perspectives on nitrogen-fixing microorganisms in tropical and subtropical oceans. *Trends in Microbiology* **8**:68-73.

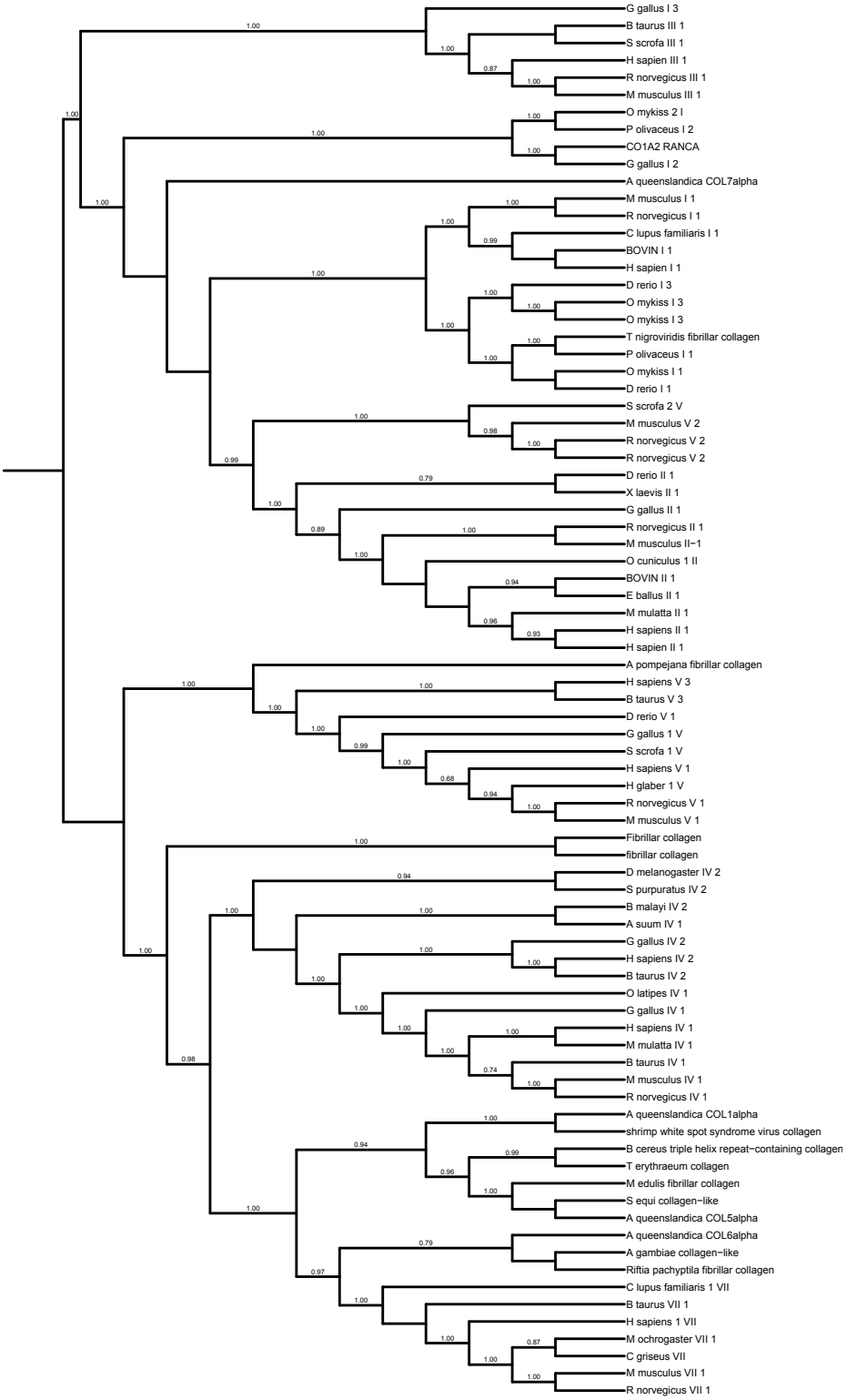
Zhang, Y. 2008. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* **9**:40.

Zhao, F., Zhang, X., Liang, C., Wu, J., Bao, Q. & Qin, S. 2006. Genome-wide analysis of restriction-modification system in unicellular and filamentous cyanobacteria. *Physiological Genomics* **24**:181-90.

Appendix A: TrpA Amino Acid Sequence

```
   1 mrkemedfld qivfpdtssf qstpintlvp gesitvpgfe lvngppiadp slfievpipi
  61 ipgspgpvgp agpsgapgpv gpigpsgapg pvgpvgpigp vgpagadgvp glagpvgpag
 121 adgvpgltgp igpigpsgap gpvgpvgptg apgpagpvgp agadgvpgla gpagpigpvg
 181 pagadgvpgl tgpigpigps gapgpvgpig pvgpagadgv pglagpvgpv gpvgptgapg
 241 pagpvgpvgp vgpagadgvp glagpvgpvg pvgptgapgp aglagpagpi gpvgpagadg
 301 vpgltgpigp igpsgapgpa gpigpvgpag adgvpglagp vgpvgpvgpv gpvgpvgpig
 361 ltgapgpagp igpvgpagad gvpgltgpig pigpsgapgp agpigpvgpa gadgvpglag
 421 pvgpvgpvgp agadgvpgla gpvgpigpvg pvgptgapgp agpvgpigpv gpagadgvpg
 481 lagpvgpvgp tgapgpagpv gpigpvgpag adgvpglagp agapgpagpv gpvgptgapg
 541 pagpagpigp vgsagadgvp gltgpigpig psgapgpagp igpvgpagad gvpglagpag
 601 asgpagpvgp igpvgpagad gvpglagpvg pvgptgapgp vgpigpvgpa gadgvpglag
 661 pagapgpagp igpvgpagtd gvpglagpvg pvgpsgapgp agpigpvgpt gapgpagpig
 721 pvgpagpagg pagpagpigp vgpagpagpi gpvgpagsdg vpgltgpigp igpsgapgpv
 781 gpigpvgsag adgvpglagp agapgpagpi gpvgpagadg vpgltgpigp igpsgapgpv
 841 gpvgpvgptg apgpvgpvgp vgpigltgap gpagpigpvg pagadgvpgl agpvgpvgpv
 901 gptgapgpag pagpigpvgp agadgvpglt gpigpigpsg apgpvgpigp vgpagadgvp
 961 glagpvgpvg pvgpvgptga pgpagpagpi gpvgpagadg vpgltgpigp igpsgapgpa
1021 gpvgpigpvg pagadgvpgl agpvgpvgpt gapgpvgpvg pvgpigltga pgpagpigpv
1081 gpagadgvpg lagpvgpvgp vgptgapgpa gpagpigpvg pagadgvpgl tgpigpigps
1141 gapgpvgpig pvgpagadgv pglagpvgpv gpvgpvgpvg ptgapgpagp agpigpvgpa
1201 gadgvpgltg pigpigpsga pgpagpvgpi gpvgpagadg vpglagpvgp vgptgapgpv
1261 gpvgpvgpig ltgapgpagp igpvgpagad gvpgltgpig pigpsgapgp agpigpvgpa
1321 gadgvpglag pvgpvgpaga pgpagpigpv gpagadgvpg lagpagapgp agpvgpigpv
1381 gpagadgvpg lagpvgpigp vgpvgpvgsa gadgvpglag pagpvgpigp vgpagadgvp
1441 glagpagpig pvgptgapgp igpigpsgap gpvgpigpvg pagadgipgl agpvgpvgpv
1501 gptgapgpig pigpsgapgp vgpigpvgpa gadgvpglag pagapgpagp igpvgpagad
1561 gvpgltgpig pigpsgapgp vgpvgpvgpt gapgpvgpig pvgpagadgv pglagpagap
1621 gpagpigpvg pagapgpagp igpvgpigps gapgpvgpvg pagadgapgf tgpvgpvgpa
1681 gpigatgpvg pagadgapgf tgpvgpvgpa gpigatgpvg ptgpvgpagp vgptgpigpg
1741 iptlpigyii agdnsptdfi tgeivlwpgl svppgffsle gqvlpitgnn elfnivgnqy
1801 ggdgvnnfvl pnlvgispvg rqlnqiedsv tgqlfvefps gsgqfinadd fpgiaggntp
1861 lgqnlnnniv vpa
```

Appendix B: Phylogenetic Analysis with Species



Vita

Simara Price
Date of Birth: August 23, 1986
Place of Birth: Brooklyn, NY, USA

Education

Doctor of Philosophy in Biology, Drexel University, Philadelphia	May 2013
Bachelor of Science in Biology, University of Maryland, College Park	May 2008

Research Experience

- Research Assistant, Drexel University — Sept. 2008-Present
- Undergraduate Research Assistant, University of Maryland — Jan. 2008-May 2008

Teaching Experience

- Teaching Assistant, Drexel University — March 2012-Present
- Teaching Assistant, Drexel University — Sept. 2008-Jan. 2010
- Undergraduate Teaching Assistant, University of Maryland — Jan. 2008-May 2008

Publications

1. Price, Simara and Anandan, Shivanthi. Characterization of a novel collagen-like protein in the cyanobacterium *Trichodesmium erythraeum. J. Phycol* (Accepted)
2. Price, Simara, Toal, Siobhan and Anandan, Shivanthi. Uncovering the structure and classification of TrpA in the cyanobacterium *Trichodesmium erythraeum* (In Preparation)

Fellowships and Awards

- American Society for Microbiology and Burroughs Wellcome Fund Science Teaching Fellows Program — Dec. 2012-Present

Professional Affiliations and Memberships

- American Association for the Advancement of Science — Sept. 2011-Present
- Drexel Graduate Women in Science and Engineering — Sept. 2010-Present
- Biology Graduate Students Association — Sept. 2008-Present
- American Society for Microbiology — Dec. 2008-Present